Edinburgh Research Explorer

KiDS-1000 Methodology

KiDS-1000 Methodology: Modelling and inference for joint weak gravitational lensing and spectroscopic galaxy clustering analysis

B. Joachimi[1], C.-A. Lin[2], M. Asgari[2], T. Tröster[2], C. Heymans[2, 3], H. Hildebrandt[3], F. Köhlinger[3], A. G. Sánchez[4], A. H. Wright[3], M. Bilicki[5], C. Blake[6], J. L. van den Busch[3], M. Crocce[7, 8], A. Dvornik[3], T. Erben[9], F. Getman[10], B. Giblin[2], H. Hoekstra[11], A. Kannawadi[12, 11], K. Kuijken[11], N. R. Napolitano[13], P. Schneider[9], R. Scoccimarro[14], E. Sellentin[11], H. Y. Shan[15, 16], M. von Wietersheim-Kramsta[1], and J. Zuntz[2]

[1] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
e-mail: b.joachimi@ucl.ac.uk
[2] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK
e-mail: calin@roe.ac.uk
[3] Ruhr-Universität Bochum, Astronomisches Institut, German Centre for Cosmological Lensing (GCCL), Universitätsstr. 150, 44801 Bochum, Germany
[4] Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, Gießenbachstr., 85741 Garching, Germany
[5] Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668, Warsaw, Poland
[6] Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia
[7] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[8] Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain
[9] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
[10] INAF - Astronomical Observatory of Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
[11] Leiden Observatory, Leiden University, P.O.Box 9513, 2300RA Leiden, The Netherlands
[12] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA
[13] School of Physics and Astronomy, Sun Yat-sen University, Guangzhou 519082, Zhuhai Campus, P.R. China
[14] Center for Cosmology and Particle Physics, Department of Physics, New York University, NY 10003, New York, USA
[15] Shanghai Astronomical Observatory (SHAO), Nandan Road 80, Shanghai 200030, China
[16] University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

We present the methodology for a joint cosmological analysis of weak gravitational lensing from the fourth data release of the ESO Kilo-Degree Survey (KiDS-1000) and galaxy clustering from the partially overlapping Baryon Oscillation Spectroscopic Survey (BOSS) and the 2-degree Field Lensing Survey (2dFLenS). Cross-correlations between BOSS and 2dFLenS galaxy positions and source galaxy ellipticities have been incorporated into the analysis, necessitating the development of a hybrid model of non-linear scales that blends perturbative and non-perturbative approaches, and an assessment of signal contributions by astrophysical effects. All weak lensing signals were measured consistently via Fourier-space statistics that are insensitive to the survey mask and display low levels of mode mixing. The calibration of photometric redshift distributions and multiplicative gravitational shear bias has been updated, and a more complete tally of residual calibration uncertainties was propagated into the likelihood. A dedicated suite of more than 20 000 mocks was used to assess the performance of covariance models and to quantify the impact of survey geometry and spatial variations of survey depth on signals and their errors. The sampling distributions for the likelihood and the χ^2 goodness-of-fit statistic have been validated, with proposed changes for calculating the effective number of degrees of freedom. The prior volume was explicitly mapped, and a more conservative, wide top-hat prior on the key structure growth parameter $S_8 = \sigma_8 \, (\Omega_m/0.3)^{1/2}$ was introduced. The prevalent custom of reporting S_8 weak lensing constraints via point estimates derived from its marginal posterior is highlighted to be easily misinterpreted as yielding systematically low values of S_8, and an alternative estimator and associated credible interval are proposed. Known systematic effects pertaining to weak lensing modelling and inference are shown to bias S_8 by no more than 0.1 standard deviations, with the caveat that no conclusive validation data exist for models of intrinsic galaxy alignments. Compared to the previous KiDS analyses, S_8 constraints are expected to improve by 20 % for weak lensing alone and by 29 % for the joint analysis.

Key words. Cosmological parameters – Gravitational lensing: weak – Large-scale structure of Universe – Methods: data analysis – Methods: analytical – Methods: statistical

Contents

1. Introduction

The Λ-cold dark matter (ΛCDM) concordance model of cosmology remains a resounding success, explaining a plethora of observations with little more than a handful of free parameters. Without alternative theories on the horizon that come anywhere close to rivalling ΛCDM in terms of completeness and predictiveness, the focus is turning increasingly towards nascent discrepancies between cosmological probes when interpreted within the standard model. This approach is fruitful: we are either witnessing the first hints towards new physics that will eventually lead to a revision or extension of ΛCDM, or we have encountered an important and persistent limitation in the fidelity of one or more key cosmological probes that is critical to push back for forthcoming cosmological experiments to succeed.

While tension in the value of the Hubble constant H_0 as inferred from local probes and early-Universe physics in combination with fluctuations in the cosmic microwave background (CMB) has received the most widespread attention to date (e.g. Verde et al. 2019), a similarly puzzling systematic difference between CMB and low-redshift measurements of the amplitude of matter density fluctuations has persisted for at least the past seven years (cf. Heymans et al. 2013), albeit at a lower significance. CMB constraints on the amplitude of the primordial matter power spectrum can be extrapolated to redshift zero assuming standard model structure growth. Alternatively, low-redshift probes of the cosmic large-scale structure provide a more direct measure of the growth amplitude, usually expressed in terms of the standard deviation of matter density fluctuations in spheres of radius $8\,h^{-1}$ Mpc (where $H_0 = 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$), denoted by σ_8.

Cosmic shear, the coherent distortions of faint galaxy images through weak gravitational lensing by the intervening large-scale matter distribution, is highly sensitive to the strength of density fluctuations projected along the line of sight, as well as ratios of angular diameter distances between observer, lenses, and sources. Consequently, it primarily constrains a degenerate combination of σ_8 and the density parameter of matter (both dark and baryonic matter), Ω_{m}, by convention usually expressed via the derived parameter $S_8 = \sigma_8\,(\Omega_{\mathrm{m}}/0.3)^{1/2}$. The most recent, final analysis of the ESA *Planck* mission produced a constraint of $S_8 = 0.83 \pm 0.02$ (68 % credible region) from its primary CMB probes (Planck Collaboration et al. 2020a,b).

By contrast, recent cosmic shear measurements resulted in $S_8 = 0.74 \pm 0.04$ for the first $450\,\mathrm{deg}^2$ of data from the ESO Kilo-Degree Survey (KiDS; Hildebrandt et al. 2020a, KV450 hereafter) and $S_8 = 0.79^{+0.02}_{-0.03}$ for the first year of data from the Dark Energy Survey (DES; Troxel et al. 2018b). Early data from the Hyper-SuprimeCam (HSC) Survey yielded $S_8 = 0.78 \pm 0.03$ (Hikage et al. 2019). At least part of the variation between the weak lensing-derived S_8 values can be attributed to the differences in the calibration of galaxy redshift distributions, with the central value of the DES analysis decreasing by 0.03 when switching to the approach taken by KiDS (Joudaki et al. 2020), and the KV450 S_8 best-fit decreasing by 0.02 when selecting galaxy samples that allow for robust calibration (Wright et al. 2020b). Independently of the calibration approach taken, however, the three concurrent surveys are in good agreement, with DES and KiDS in combination reaching $\sim 3\sigma$ tension with *Planck* (Asgari et al. 2020c; Joudaki et al. 2020). Intriguingly, fully independent measurements of the full redshift-space power spectrum of spectroscopic galaxy clustering with the Baryon Oscillation Spectroscopic Survey (BOSS) show the same tendency towards lower values of S_8 (d'Amico et al. 2020; Colas et al. 2020; Ivanov et al. 2020; Tröster et al. 2020b).

While cosmic shear is attractive as a direct probe of the dark and bright matter distribution with few assumptions about the nature or state of this matter, its statistical power derives mainly from non-linear scales that cannot be modelled analytically from first principles. The most recent cosmic shear analyses have already reduced the statistical errors to the same level as residual systematic uncertainty caused by both astrophysical and measurement biases that need to be carefully removed or modelled (see Mandelbaum 2018 for a recent review). The joint analysis of cosmic shear with galaxy clustering in overlapping sky areas, including the cross-correlations between galaxy positions and gravitational shear in the background commonly known as galaxy-galaxy lensing (GGL), offers additional cosmological constraining power and the removal of parameter degeneracies,

as well as self-calibration of some of the systematics (Bernstein 2009; Joachimi & Bridle 2010). Therefore, the combination of cosmic shear with foreground samples of galaxy 'lenses' is being established as the default analysis approach to extracting gravitational lensing information from large-scale structure (van Uitert et al. 2018; Joudaki et al. 2018; Abbott et al. 2018).

This work details the approach to the joint probes modelling and likelihood analysis of Data Release 4 (DR4; Kuijken et al. 2019) of KiDS. The data set will be referred to as KiDS-1000, covering a raw survey area of $1006 \deg^2$ out of the final KiDS area of $1350 \deg^2$. The KiDS footprint has near-complete coverage by spectroscopic galaxy surveys – BOSS (Eisenstein et al. 2011) in the north and the 2-degree Field Lensing Survey (2dFLenS; Blake et al. 2016) in the south. The latter is designed to provide similar galaxy samples to BOSS albeit with sparser sampling. We opt to combine KiDS-1000 cosmic shear with GGL in the BOSS and 2dFLenS overlap and an existing galaxy clustering analysis over the full BOSS survey area by Sánchez et al. (2017, S17 hereafter). The approach is an extension of the earlier analyses by Tröster et al. (2020b) who combined S17 with cosmic shear only from KV450 and Joudaki et al. (2018) who combined the first KiDS cosmic shear results (Hildebrandt et al. 2017) with clustering and GGL measured in 2dFLenS and BOSS. An analogous investigation for DES was carried out by Krause et al. (2017), where both lens and source samples were derived from the same survey.

The paper is structured as follows: Section 2 presents our choice of observables and the model to jointly predict the corresponding cosmological and astrophysical signals. Section 3 summarises the characteristics and processing of our data sets and details the calibration and modelling of measurement systematics. An extensive suite of survey simulations is introduced in Sect. 4, and the impact of survey geometry and the spatial variation of survey characteristics on the cosmological signals is assessed. In Sect. 5 a detailed discussion of our approach to modelling the joint covariance of the observables, and its accuracy, is provided. Section 6 covers the choice of model parameters and the form of the likelihood, as well as how we report goodness of fit and parameter constraints in the context of large, highly non-Gaussian parameter spaces. The analysis pipeline and associated modelling choices are validated in Sect. 7 before concluding in Sect. 8.

Companion papers present the construction and calibration of gravitational shear catalogues (Giblin et al. 2020) and redshift distributions (Hildebrandt et al. 2020b), the cosmological analysis of cosmic shear (Asgari et al. 2020b) and of joint clustering and weak lensing (Heymans et al. 2020), and an investigation of extensions to the spatially flat ΛCDM model considered throughout here (Tröster et al. 2020a). Unless otherwise specified, calculations and plots use the fiducial values of our set of model parameters specified in Table A.1.

2. Signal modelling

Including non-linear scales is critical to unlock the full constraining power of large-scale structure cosmological probes. Traditionally, galaxy clustering analyses have opted for perturbative approaches to modelling signals into the mildly non-linear regime to wavenumbers of $k \sim 0.3 \, h \, \text{Mpc}^{-1}$, which enables closed-form and consistent expressions for non-linear structure growth, galaxy bias, and redshift-space effects (see Desjacques et al. 2018 for a review). Cosmic shear, in contrast, routinely accesses scales at $k > 1 \, h \, \text{Mpc}^{-1}$, where its signal-to-noise is high, via empirical models based on the halo model paradigm

that have been calibrated on N-body and/or full hydro-dynamical simulations (beginning with Hamilton et al. 1991; see also the seminal work by Smith et al. 2003). This makes the consistent modelling of a joint clustering and cosmic shear analysis, especially for the GGL cross-correlation, a major challenge for current and forthcoming analyses (e.g. Bose et al. 2020). We choose a hybrid approach that blends models previously used successfully by the BOSS and KiDS teams, as detailed in Sects. 2.1 and 2.2.

The second major topic of this section is the choice of two-point statistics used in the analysis. As we demonstrate, the clustering and weak lensing (i.e. cosmic shear and GGL) probes can be treated independently, so that we can simply adopt the original statistics employed in S17, that is to say correlation functions as a function of three-dimensional separation and the angle subtended with the line of sight, binned into wedges. For GGL and cosmic shear we opt for band powers derived from angular correlation functions, closely following van Uitert et al. (2018). They are straightforward to compute and combine good scale localisation in both configuration and Fourier space, which helps to separate interpretable cosmological information from contamination by systematics and astrophysical signals (cf. Asgari et al. 2020a; Park et al. 2020 for recent examples in GGL). Examples include additive shear bias on large configuration-space scales, blending of galaxy images on small configuration-space scales, baryonic effects on the matter power spectrum at high angular frequencies, and computationally expensive curved-sky expressions at low angular frequencies.

While the majority of weak lensing analyses still rely on correlation functions, which have sensitivity to a very broad range of Fourier modes of the matter distribution (e.g. Asgari et al. 2020c), some direct Fourier-space measures like pseudo-Cls (Hikage et al. 2019) and quadratic maximum likelihood estimators (Köhlinger et al. 2017) have seen recent applications. Becker et al. (2016) calculated both pseudo-Cl and band power spectra, but only used real-space measures in their main likelihood analysis, as did the most recent cosmic shear papers from KiDS and DES. Correlation function-derived band powers have two additional advantages over direct Fourier-space statistics in that they are not sensitive to the survey geometry, and in that they do not require an estimate of the noise power, which is trivially removed by excluding zero-lag correlation functions in the band power computation. A detailed comparison between correlation functions, band powers, and compressed statistics with exact separation of E- and B-modes (versus the approximate separation offered by band powers) is presented in Asgari et al. (2020b).

2.1. Matter power spectra

Linear matter power spectra $P_{\text{m,lin}}(k, z)$ as a function of wavenumber k and redshift z are calculated with Camb[1] (Lewis et al. 2000; Lewis & Bridle 2002; Howlett et al. 2012). Massive standard-model neutrinos are included assuming the normal hierarchy at the minimum sum of masses, $\sum m_\nu = 0.06 \, \text{eV}$. We keep this parameter fixed in our default setup, in line with Planck Collaboration et al. (2020a), and consider constraints on neutrino mass in a companion paper (Tröster et al. 2020a).

We derive two non-linear matter power spectra from the linear one. First, we compute a non-perturbative model using the Mead et al. (2015) HMCode included in Camb, whose validity

---

[1] Code for Anisotropies in the Microwave Background; https://camb.info

extends into the deeply non-linear regime, recovering N-body results at the few per-cent level to $k = 10\,h\,\mathrm{Mpc}^{-1}$. The halo model approach in HMCODE incorporates baryonic feedback (primarily through Active Galactic Nuclei; AGN) on the matter distribution via a halo bloating parameter η_0 and the amplitude of the halo mass-concentration relation, denoted by A_{bary}. Following Joudaki et al. (2018), we fix the relation between the two feedback parameters to $\eta_0 = 0.98 - 0.12 A_{\mathrm{bary}}$ and only use A_{bary} as a free parameter in our analysis. The resulting matter power spectrum is denoted by $P_{\mathrm{m,nl}}$.

The choice of a one-parameter feedback model is justified as, on relevant spatial scales, it captures the range of matter power spectrum models extracted from hydro-dynamic simulation which are consistent with observations (see Chisari et al. 2019 for a review). KiDS analyses have only very weak sensitivity to the scale substantially affected by baryon feedback as evidenced by the unconstrained posterior on A_{bary} (Hildebrandt et al. 2020a); this remains the case for KiDS-1000 (Asgari et al. 2020b). More conservative modelling choices, including a parametrisation of the possible ranges of matter power spectrum modifications due to feedback via principal component analysis (Eifler et al. 2015), were investigated by Huang et al. (2019). For scales slightly less conservative than those used in KiDS-1000, they found minimal impact on S_8 and moderate shifts along the $\Omega_{\mathrm{m}} - \sigma_8$ degeneracy, which are captured well by A_{bary}.

The second, perturbative non-linear power spectrum, $P_{\mathrm{m,nl-pt}}$, is calculated using renormalised perturbation theory (RPT; Crocce & Scoccimarro 2006) to one-loop order. We adopt the updated version applied in S17 (see also Eggemeier et al. 2020), dubbed gRPT, that incorporates a re-summation of the mode coupling terms consistently with the propagator ones (i.e. those proportional to the linear power spectrum) to make the theory Galilean invariant (see also Taruya et al. 2012). S17 demonstrated with N-body simulations that this model is accurate to 2 % or better out to scales of at least $k = 0.25\,h\,\mathrm{Mpc}^{-1}$.

We compare the two non-linear power spectrum models with the alternative dark-matter only Halofit prescription by Takahashi et al. (2012), as well as the CosmicEmu emulated power spectrum by Heitmann et al. (2014) in the top two panels of Fig. 1. The comparison is at $z = 0.38$, which is the mean redshift of one of our two 'lens' galaxy samples. We treat CosmicEmu as the truth in this case although the emulator itself is limited to 2 % accuracy on relevant non-linear scales. All non-linear recipes capture the deviation from the linear matter power spectrum well, with the perturbative model starting to break down beyond $k = 0.25\,h\,\mathrm{Mpc}^{-1}$, in line with the S17 results. Our default Mead et al. (2015) model is systematically higher than CosmicEmu in the mildly non-linear regime[2] around $k \approx 0.2\,h\,\mathrm{Mpc}^{-1}$, while Halofit shows a similar trend around $k = 1\,h\,\mathrm{Mpc}^{-1}$. Systematic trends at high k are to some degree absorbed by the freedom in the baryon feedback amplitude, with the allowed prior range shown as the shaded region in Fig. 1. However, this would not be the case for the excess power of the Takahashi et al. (2012) Halofit model as we limit the feedback model to only account for suppression of power.

Recently, Euclid Collaboration et al. (2019) introduced the EuclidEmulator and demonstrated it to be in excellent agreement with CosmicEmu to $\sim 1\%$ at $z \lesssim 1$ and over the majority of relevant k-scales, despite using completely independent simulations
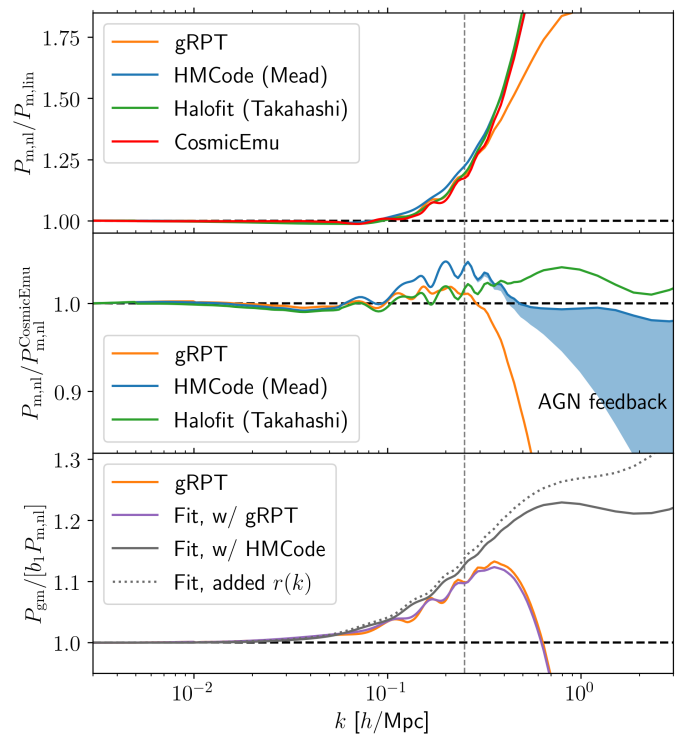
Fig. 1. Comparison of 3D power spectra, computed at $z = 0.38$ and the parameters listed in Table A.1. *Top*: Non-linear matter power spectra relative to the linear matter power spectrum, for the gRPT perturbative model, the Mead et al. (2015) HMCODE, the Takahashi et al. (2012) Halofit prescription, and the CosmicEmu emulated power spectrum (Heitmann et al. 2014). *Centre*: Non-linear matter power spectra relative to the CosmicEmu model. The blue shaded region covers the power spectrum range within the prior range of our AGN feedback description. *Bottom*: Full galaxy-matter power spectrum P_{gm} relative to the Mead et al. (2015) power spectrum multiplied by the linear galaxy bias, shown for the gRPT perturbation theory model (orange) and the fit formula of Eq. (3), using gRPT (purple) or Mead et al. (2015) (grey) for the non-linear matter power spectrum term. The dotted grey curve includes an additional $r(k)$ term obtained from a semi-analytic model of a galaxy sample similar to the one used in our analysis. The vertical grey line indicates the smallest scales used in the galaxy clustering modelling.

and emulation methods[3]. The emulators agree that the Takahashi et al. (2012) Halofit model consistently over-predicts power over a broad range of redshifts and cosmological parameters by typically 3 % at $k \sim 1\,h\,\mathrm{Mpc}^{-1}$ (Heitmann et al. 2014; Mead et al. 2015; Euclid Collaboration et al. 2019). We can therefore reject Halofit as an accurate description of matter power at highly non-linear scales.

Nevertheless, it is relevant to include Halofit in comparisons as it is still widely used, for instance in the recent DES Year 1 (Troxel et al. 2018b) and HSC (Hikage et al. 2019) cosmic shear analyses. However, Troxel et al. (2018b) discarded highly non-linear scales that are affected by baryonic feedback at 2 % or more, which also suppresses potential inaccuracies in matter power spectrum modelling to negligible levels. Hikage et al.

---

[2] We note that this excess is parameter-dependent. For different cosmological parameter values deviations can also be negative, but they tend to be at a similar level and on the same scales.

[3] We cannot use either emulator in the main analysis because they do not include baryonic feedback. Furthermore, CosmicEmu is limited to a maximum redshift of $z = 2$, while the real-data redshift distributions extend to higher redshifts. The EuclidEmulator only covers a restricted parameter space, e.g. it does not model our fiducial value of σ_8, so that it cannot be added to Fig. 1.

(2019) repeated their analysis with HMCode in a similar setup to ours and found a tolerable 0.2σ shift in the parameter combination $\sigma_8 (\Omega_{\rm m}/0.3)^{0.45}$.

To assess the impact of modelling inaccuracies in the matter power spectrum on KiDS-1000, we perform a mock likelihood analysis of the joint clustering and weak lensing probes with our default HMCode model (for details see Sect. 7), constructing a noiseless mock data vector with (1) CosmicEmu and (2) Takahashi et al. (2012) Halofit as the non-perturbative non-linear matter power spectrum. This results in a shift of the best-fit S_8 from the input value by 0.1σ for CosmicEmu, with the best-fit $A_{\rm bary}$ very close to the dark-matter only value of 3.13. This demonstrates that our HMCode model accurately captures the state-of-the-art emulator predictions. The corresponding S_8 shift in the case of Halofit is 0.3σ; if we had used this as our model, it would have rendered the non-linear matter power spectrum prescription the dominant systematic in KiDS cosmological modelling and inference.

2.2. Galaxy power spectra

The galaxy power spectrum $P_{\rm gg}(k, z)$ underlying the galaxy clustering signals is adopted from S17 and based on the perturbation theory approach developed by Chan et al. (2012). It is of the form

$$P_{\rm gg}(k, z) = \sum_{\alpha,\beta} \alpha\beta \, P_{\alpha\beta}(k, z) + b_1 \gamma_3^- \, P_{b_1\gamma_3^-}(k, z) + P_{\rm noise}(k, z) , \quad (1)$$

with $\alpha, \beta \in [b_1, b_2, \gamma_2]$. We have introduced the linear and quadratic bias parameters b_1 and b_2, and the non-local bias parameters γ_2 and γ_3^-. The different power spectrum terms on the right-hand side of Eq. (1) are convolutions of the linear power spectrum and given explicitly in Appendix A of S17, with the identification $P_{b_1 b_1} \equiv P_{\rm m,nl-pt}$.

In the same formalism the cross-power spectrum between the galaxy and matter distribution required for GGL is given by

$$P_{\rm gm,pt}(k, z) = b_1 P_{\rm m,nl-pt}(k, z) + b_2 P_{b_2}(k, z) + \gamma_2 P_{\gamma_2}(k, z) \quad (2)$$
$$+ \gamma_3^- P_{\gamma_3^-}(k, z) ,$$

where the right-hand side power spectrum terms are again provided in S17. To extend the model to smaller scales while ensuring consistency with the non-linear model for the cosmic shear signal, we undertake the following empirical modifications to this model: first, we replace the perturbative non-linear power spectrum in the linear bias term with the halo model-based version, $P_{\rm m,nl}$. We also add an additional function $r(k)$ that will be discussed further below.

Secondly, we re-formulate the remaining terms relative to the square of the linear matter power spectrum, which results in smoothly varying functions $\mathcal{F}$ that are robustly extrapolated into the deeply non-linear regime. To a very good approximation, these functions can also be assumed redshift-independent (an approach also adopted by S17). This leads to the following model:

$$P_{\rm gm}(k, z) = b_1 r(k) P_{\rm m,nl}(k, z) + \left\{ b_2 \, \mathcal{F}_{b_2}(k) - \gamma_2 \, \mathcal{F}_{\gamma_2}(k) \right. \quad (3)$$
$$\left. - \gamma_3^- \, \mathcal{F}_{\gamma_3^-}(k) \right\} P_{\rm m,lin}^2(k, z) ,$$

with the definition

$$\mathcal{F}_\alpha(k) := \frac{P_\alpha(k, z_{\rm ref})}{P_{\rm m,lin}^2(k, z_{\rm ref})} = \frac{h^3}{\rm Mpc^3} \exp \left\{ \sum_{i=0}^{2} f_{\alpha,i} \left[\ln \left(k \frac{\rm Mpc}{h} \right) \right]^i \right\}$$

$$(4)$$

for each higher-order bias term, that is $\alpha \in [b_2, \gamma_2, \gamma_3^-]$. We express the logarithm of the functions $\mathcal{F}$ as second-order polynomials in $\ln k$ and fit the coefficients $f_{\alpha,i}$ to the perturbation theory model within its range of validity. We find that including quadratic terms in Eq. (4) is sufficient to capture the deviations in scale dependence from $P_{\rm m,lin}^2$. Since the non-local bias terms in Eq. (2) are consistently negative, we extract an overall minus sign and model the absolute values in $\mathcal{F}_{\gamma_2}$ and $\mathcal{F}_{\gamma_3^-}$.

Most of the cosmology dependence of the power spectrum terms in Eq. (3) is captured by the linear matter power spectrum. Due to the convolutions in these terms there is some mode mixing which induces sensitivity to cosmological parameters that modify the shape of the power spectrum. We find it is sufficient to account for $n_{\rm s}$, the power-law slope of the primordial power spectrum, and the combination $\omega_{\rm c}/h$, where $\omega_{\rm c} = \Omega_{\rm c} h^2$ is the physical cold dark matter density. The latter governs the position of the peak in the matter power spectrum (strictly speaking, $\Omega_{\rm m} h$ should capture this dependence even better, but we find using the baseline parameter $\omega_{\rm c}$ sufficient). Again, a polynomial including terms up to second order is sufficient to capture the dependence, yielding

$$f_{\alpha,i} := \sum_{m,n=0}^{m+n\leq 2} g_{\alpha,i}^{mn} \left(\frac{\omega_{\rm c}}{h} \right)^m n_{\rm s}^n . \quad (5)$$

The complete set of best-fit coefficients $g_{\alpha,i}^{mn}$ is provided in Table A.2. We note that the ansatz of Eq. (3) has the additional benefit of a speed-up over the perturbative calculations by several orders of magnitude in some parts of parameter space, removing a major bottleneck in the weak lensing likelihood evaluations.

Non-linear galaxy bias and stochasticity, mostly driven by satellite galaxies, both cause a de-correlation between the matter and galaxy density fields (Dekel & Lahav 1999; Cacciato et al. 2012). This can be expressed in the form of deviations of the correlation coefficient $r(k, z) = P_{\rm gm}(k, z)/\sqrt{P_{\rm gg}(k, z) P_{\rm m}(k, z)}$ from unity. For $k < 0.25 \, h \, {\rm Mpc}^{-1}$, gRPT predicts r to be very close to 1, in good agreement with recent observational and simulation studies (Blake et al. 2011; Simon & Hilbert 2018; Dvornik et al. 2018; see also Asgari et al. 2020a). However, these works also see a more pronounced deviation from $r = 1$ beginning on scales just below the maximum k that can be modelled perturbatively. Therefore, we have considered including an explicit function $r(k)$ into the first term of Eq. (3) to enable additional flexibility in galaxy biasing on scales beyond where perturbation theory is valid. For illustration we adopt the functional form presented in Asgari et al. (2020a), which was derived from semi-analytic modelling in N-body simulations for a BOSS CMASS-like sample (as originally developed in Simon & Hilbert 2018).

The bottom panel of Fig. 1 illustrates the non-linear bias modelling in $P_{\rm gm}$. gRPT predicts an increase in power of up to 10 % relative to linear galaxy bias alone at $k = 0.25 \, h \, {\rm Mpc}^{-1}$. Its scale dependence is accurately reproduced by the fit formula of Eq. (3) when using the gRPT non-linear matter power spectrum for the b_1 term. Switching to the Mead et al. (2015) matter power spectrum mildly increases power with respect to gRPT where the latter is valid, in line with the small excess seen in $P_{\rm m,nl}$. The non-perturbative matter power spectrum smoothly extends the non-linear biasing trend to smaller scales, with the excess power levelling off above $k \approx 0.5 \, h \, {\rm Mpc}^{-1}$. Incorporating the Asgari et al. (2020a) expression for $r(k)$ enhances power significantly further beyond those scales.

Since realistic simulations to highly non-linear scales of both the lens and source galaxy samples are currently not available to us for validation, we consider the approach to galaxy biasing in the highly non-linear regime outlined here as currently too speculative to adopt in our cosmological analysis. Therefore, we set $r \equiv 1$ and limit the GGL signals to scales that only have significant contributions from $k < 0.3\,h\,\mathrm{Mpc}^{-1}$, where the perturbative approach holds and our empirical modifications are minor. While we thus discard high signal-to-noise GGL measurements, the impact on the overall constraining power of our data sets is minimal, largely because we make use of clustering signals measured over a much larger survey area than GGL (see Sect. 3). We note however that forthcoming analyses where the clustering and GGL areas are compatible would suffer from such a conservative approach, making the development of a robust non-linear biasing model a priority. The hybrid approach outlined above, or a more comprehensive halo model of all astrophysical contributions to the large-scale structure probes involved (see e.g. Fortuna et al. 2020), are promising Ansätze in this regard.

2.3. Clustering summary statistics

The translation from three-dimensional galaxy power spectrum to the clustering observable follows S17, with the key steps summarised here. As a first step, redshift-space effects are taken into account. If μ denotes the cosine of the angle between the separation of galaxy pairs and the line of sight, the redshift-space power spectrum is given by

$$P_{\mathrm{gg,s}}(k,\mu,z) = W_\infty(f_g\,k\,\mu)\left\{P_{\mathrm{gg}}(k,z) + 2f_g(a)\,\mu^2 P_{\mathrm{g}\theta}(k,z)\right. \tag{6}$$
$$\left. + f_g(a)^2\mu^4 P_{\theta\theta}(k,z) + \text{higher-order terms}\right\},$$

where P_{gg} is taken from Eq. (1) and

$$f_g(a) = \frac{\mathrm{d}\ln D(a)}{\mathrm{d}\ln a} \tag{7}$$

is the structure growth rate, with $D(a)$ the linear growth factor and scale factor $a = (1+z)^{-1}$. The galaxy-velocity cross-power spectrum reads

$$P_{\mathrm{g}\theta}(k,z) = b_1 P_{\mathrm{m}\theta}(k,z) + b_2\,P_{b_2}(k,z) + \gamma_2\,P_{\gamma_2}(k,z) + \gamma_3^-\,P_{\gamma_3^-}(k,z), \tag{8}$$

where $P_{\mathrm{m}\theta}(k,z)$ and $P_{\theta\theta}(k,z)$ in Eq. (6) are the matter density contrast cross-power spectrum, and the auto-power spectrum, of the velocity divergence θ, respectively[4]. These terms are also derived from gRPT, using the same prescription as for the density spectra (Crocce et al. 2012), and do not contain additional dependencies on our bias parameters. The remaining power spectrum terms match those in Eq. (2). We refer the reader to S17 for the explicit form of the higher-order terms in Eq. (6). Fingers of god generated by small-scale virial motion are modelled by the term

$$W_\infty(x) = \frac{1}{\sqrt{1+x^2 a_{\mathrm{vir}}^2}}\,\exp\left(-\frac{x^2\sigma_v^2}{1+x^2 a_{\mathrm{vir}}^2}\right), \tag{9}$$

where σ_v is the one-dimensional linear velocity dispersion, which is calculated from $P_{\theta\theta}$ (cf. Eq. 39 in Scoccimarro 2004).

The quantity a_{vir} is a free parameter in the analysis that accounts for small-scale, non-linear contributions to the velocities and therefore deviations from Gaussianity in the velocity distribution.

The redshift-space galaxy power spectrum is transformed to correlation functions as a function of comoving lag s, μ, and redshift z as follows,

$$\xi_{\mathrm{gg}}(s,\mu,z) = \sum_{l=0}^{2} L_{2l}(\mu)\frac{(-1)^l(4l+1)}{(2\pi)^2}\int_0^\infty \mathrm{d}k\,k^2\,j_{2l}(ks) \tag{10}$$
$$\times \int_{-1}^{1}\mathrm{d}\mu_1 L_{2l}(\mu_1)P_{\mathrm{gg,s}}(k,\mu_1,z),$$

where L_i denotes the Legendre polynomial of degree i, and j_i is the spherical Bessel function of order i. The expansion in Legendre polynomials is truncated at $l = 2$ as higher-order contributions are small on the scales considered[5]. The correlation function is then cast into three redshift-space wedges, equidistant in μ. The three-dimensional galaxy pair separation and μ are calculated from the observed redshift assuming a fiducial cosmology, which leads to Alcock-Paczynski distortions (Alcock & Paczynski 1979). The final prediction for the clustering observable is hence

$$\xi'_{\mathrm{gg},i}(s',z) = \frac{1}{\mu'_{i,\mathrm{up}}-\mu'_{i,\mathrm{lo}}}\int_{\mu'_{i,\mathrm{lo}}}^{\mu'_{i,\mathrm{up}}}\mathrm{d}\mu'\xi_{\mathrm{gg}}\left(s\left[\mu',s'\right],\mu\left[\mu',s'\right],z\right) \tag{11}$$

for wedges between μ_{lo} and μ_{up} and indexed by i, where primes denote quantities evaluated in the fiducial cosmology. The translation for s and μ is carried out according to

$$s_\perp = \frac{f_K(\chi[z_{\mathrm{mean}}])}{f'_K(\chi[z_{\mathrm{mean}}])}\,s'_\perp\;;\quad s_\parallel = \frac{H'(z_{\mathrm{mean}})}{H(z_{\mathrm{mean}})}\,s'_\parallel \tag{12}$$

for the perpendicular and line-of-sight components of the separation vector, respectively. We use the comoving angular diameter distance f_K, the comoving (radial) distance χ, and the Hubble parameter $H(z)$ evaluated at the mean redshifts z_{mean} of the lens galaxy samples.

As in S17, we divide the BOSS galaxies into two redshift bins, each modelled with an individual set of galaxy bias and other non-cosmological parameters; cf. Table A.1. The correlation functions overall are modelled at the mean redshift in each bin, neglecting evolution within the bin. S17 showed this approximation to be accurate on simulations that include redshift evolution. Moreover, an analysis of the same data set with angular correlation functions measured in thin redshift shells produced fully consistent results (Salazar-Albornoz et al. 2017). Overall, a likelihood analysis of mock BOSS clustering signals derived from an N-body simulation revealed subdominant biases of the posterior maximum with respect to the input cosmology of less than 0.3σ for S_8 and less than 0.4σ for Ω_{m} (S17; Tröster et al. 2020b).

2.4. Cosmic shear summary statistics

Cosmic shear two-point statistics can be expressed as linear functionals of angular power spectra, which in turn are line-of-sight projections of the three-dimensional matter power spectrum. We perform tomography, that is to say we consider all

---

[4] In this subsection θ denotes velocity divergence in accordance with the perturbation theory literature, but in the remainder of the paper we use θ as a variable for angular separation.

[5] Once the Alcock-Paczynski distortions are taken into account, the final clustering wedges may have contributions from higher multipoles.

unique combinations of a number of source galaxy sub-samples, denoted by indices i and j in the following. Statistics derived from galaxy ellipticity measurements receive two contributions, one from the gravitational lensing effect (subscript 'G') and one from intrinsic alignments (subscript 'I'; referred to as IA henceforth) of galaxy ellipticities. This results in 'observed shear' power spectra

$$C_{\epsilon\epsilon}^{(ij)}(\ell) = C_{GG}^{(ij)}(\ell) + C_{GI}^{(ij)}(\ell) + C_{IG}^{(ij)}(\ell) + C_{II}^{(ij)}(\ell) \, , \qquad (13)$$

where the right-hand side power spectra are given by Limber-approximated projections of the form (Kaiser 1992; Loverde & Afshordi 2008)

$$C_{ab}^{(ij)}(\ell) = \int_0^{\chi_{hor}} d\chi \, \frac{W_a^{(i)}(\chi) \, W_b^{(j)}(\chi)}{f_K^2(\chi)} \, P_{m,nl}\left(\frac{\ell + 1/2}{f_K(\chi)}, z(\chi)\right) \, , \qquad (14)$$

with a, b $\in$ {I, G}. The integral runs over the entire line of sight to the horizon χ_{hor}. The weak lensing kernel is

$$W_G^{(i)}(\chi) = \frac{3H_0^2\Omega_m}{2\,c^2} \frac{f_K(\chi)}{a(\chi)} \int_\chi^{\chi_{hor}} d\chi' \, n_S^{(i)}(\chi') \, \frac{f_K(\chi' - \chi)}{f_K(\chi')} \, , \qquad (15)$$

where Ω_m is the total matter density parameter and $n_S^{(i)}(\chi)$ is the probability density distribution of comoving distances of galaxies in source sample i, which in practice we express in terms of redshift[6], that is $n_S(z) = n_S(\chi) \, d\chi/dz$. We choose an IA kernel that produces the so-called NLA model (Bridle & King 2007),

$$W_I^{(i)}(\chi) = -A_{IA} \left(\frac{1 + z(\chi)}{1 + z_{pivot}}\right)^{\eta_{IA}} \frac{C_1\rho_{cr}\,\Omega_m}{D\,(a[\chi])} \, n_S^{(i)}(\chi) \, , \qquad (16)$$

where $C_1\rho_{cr} \approx 0.0134$ is constant, and z_{pivot} is an arbitrary pivot which we set to 0.3 for compatibility with earlier KiDS and IA analyses (Joachimi et al. 2011). The IA kernel introduces two additional parameters: the dimensionless IA amplitude A_{IA} and additional freedom in the redshift dependence of the IA strength via η_{IA}. As there is currently no evidence for significant extra redshift evolution in our source samples (see e.g. KV450), we only consider η_{IA} in extended analyses and set $\eta_{IA} = 0$ by default.

More flexibility in the IA model to account for the still rather limited physical understanding of galaxy alignments may seem prudent (Blazek et al. 2019; Samuroff et al. 2019), but it enables residual systematics in the source redshift distributions to disguise as a physical signal, which subsequently risks biasing constraints on primary cosmological parameters (Efstathiou & Lemos 2018). This concern would not necessarily extend to IA parameters that do not modify the overall amplitude of weak lensing two-point statistics, or the scaling of their amplitude with redshift (such as additional freedom on highly non-linear scales). However, it is this property that also makes these parameters less relevant for the cosmological inference; their absence would cause a poor goodness of fit rather than biases in the cosmological parameters if the data preferred the more complex IA model.

There is growing evidence that IA model parameters pick up residual discrepancies in the scaling of cosmic shear signals with source redshifts, rather than solely the IA contribution: IA amplitudes in both the most recent KiDS and DES analyses tend to be higher than the consensus of direct IA measurements predicts (Fortuna et al. 2020). Hikage et al. (2019) do not detect

IA, but also see a dependence of the IA amplitude constraint on the choice of photometric redshift algorithm employed. Recently, Wright et al. (2020b) showed that the fiducial KV450 constraint of $A_{IA} \sim 1$ is reduced to values fluctuating around zero for a conservative selection of source galaxies with more secure redshift calibration (we adopt this approach; see Sect. 3.3 for details). In conclusion, we continue to follow the philosophy of earlier KiDS analyses to employ the simplest, most predictive IA model as long as it yields a good fit to the data.

Band powers are defined as angular averages over the angular power spectra,

$$C_{E/B,l}^{(ij)} := \frac{1}{\mathcal{N}_l} \int_0^\infty d\ell \, \ell \, S_l(\ell) \, C_{\epsilon\epsilon,E/B}^{(ij)}(\ell) \, , \qquad (17)$$

where S_l is the band power response function for an angular bin l. The normalisation is given by

$$\mathcal{N}_l = \int_0^\infty d\ell \, \ell^{-1} \, S_l(\ell) \, , \qquad (18)$$

which is designed such that the band power traces $\ell^2 C_{\epsilon\epsilon,E/B}$ at the log-centre of bin l. Here, the spin-2 gravitational shear field has been decomposed into curl-free E-modes and divergence-free B-modes. The former carry the signal of interest, that is $C_{\epsilon\epsilon,E} \equiv C_{\epsilon\epsilon}$, whereas we assume that none of our cosmological and astrophysical signals produce B-modes intrinsically, so $C_{\epsilon\epsilon,B} \equiv 0$. For both gravitational lensing and the II-term prediction in the NLA model, B-modes are only generated at second order and therefore safely negligible in KiDS (Bernardeau et al. 1997; Schneider et al. 1998; Hirata & Seljak 2004), as are other potential sources of B-modes like source redshift clustering (Schneider et al. 2002b) or exotic physics (e.g. Thomas et al. 2017). B-mode statistics are therefore used as null tests for residual systematics in the data.

The band powers are derived from the two-point correlation functions, which are related to the angular power spectra via Hankel transformations of the form

$$\xi_+^{(ij)}(\theta) = \int_0^\infty \frac{d\ell\,\ell}{2\pi} \left\{C_{\epsilon\epsilon,E}^{(ij)}(\ell) + C_{\epsilon\epsilon,B}^{(ij)}(\ell)\right\} J_0(\ell\theta) \, ; \qquad (19)$$

$$\xi_-^{(ij)}(\theta) = \int_0^\infty \frac{d\ell\,\ell}{2\pi} \left\{C_{\epsilon\epsilon,E}^{(ij)}(\ell) - C_{\epsilon\epsilon,B}^{(ij)}(\ell)\right\} J_4(\ell\theta) \, ,$$

where J_μ is a cylindrical Bessel function of the first kind of order μ. Due to the orthogonality of the Bessel functions, these equations are readily inverted to express power spectra as a functional of $\xi_\pm$. Inserting the result into Eq. (17) yields

$$C_{E/B,l}^{(ij)} = \frac{\pi}{\mathcal{N}_l} \int_0^\infty d\theta \, \theta \, T(\theta) \left\{\xi_+^{(ij)}(\theta) \, g_+^l(\theta) \pm \xi_-^{(ij)}(\theta) \, g_-^l(\theta)\right\} \, , \qquad (20)$$

where we defined the kernels

$$g_\pm^l(\theta) = \int_0^\infty d\ell \, \ell \, S_l(\ell) \, J_{0/4}(\ell\theta) \, . \qquad (21)$$

Formally, Eq. (20) is exact for the weight function $T \equiv 1$. However, in practice the correlation functions can only be measured over a finite range of angular separations, so T has to vanish outside this range. To avoid ringing in the band power kernel due to this cut-off, we apodise the angular scales entering the band

---

[6] We employ the same symbol for the comoving distance distribution and the redshift distribution to keep notation simple. They can be distinguished by their function argument.

power computation in the form of a Hann window,

$$
T(\theta) = \begin{cases} 0 \; ; \; x < x_{\text{lo}} - \dfrac{\Delta_x}{2} \\[2mm] \cos^2\left[\dfrac{\pi}{2}\dfrac{x - (x_{\text{lo}} + \Delta_x/2)}{\Delta_x}\right] \; ; \; x_{\text{lo}} - \dfrac{\Delta_x}{2} \le x < x_{\text{lo}} + \dfrac{\Delta_x}{2} \\[2mm] 1 \; ; \; x_{\text{lo}} + \dfrac{\Delta_x}{2} \le x < x_{\text{up}} - \dfrac{\Delta_x}{2} \\[2mm] \cos^2\left[\dfrac{\pi}{2}\dfrac{x - (x_{\text{up}} - \Delta_x/2)}{\Delta_x}\right] \; ; \; x_{\text{up}} - \dfrac{\Delta_x}{2} \le x < x_{\text{up}} + \dfrac{\Delta_x}{2} \\[2mm] 0 \; ; \; x \ge x_{\text{up}} + \dfrac{\Delta_x}{2} \; . \end{cases}
$$
(22)

where $x = \log\theta$ and Δ_x is the log-width of the apodisation, which is a free parameter. The apodisation at the lower and upper angular limits is centred on the scales $x_{\text{lo}} = \log\theta_{\text{lo}}$ and $x_{\text{up}} = \log\theta_{\text{up}}$, respectively. The apodisation suppresses out-of-band sensitivity in Fourier space at the price of somewhat extending the range of angular scales used in the input correlation functions at both ends.

In this work we opt for a band power response function that is a top hat between $\ell_{\text{lo},l}$ and $\ell_{\text{up},l}$ (for other choices, see Becker & Rozo 2016). The normalisation is then given by $\mathcal{N}_l = \ln\left(\ell_{\text{up},i}/\ell_{\text{lo},i}\right)$. The functions $g_\pm$ can be expressed in closed form as (Schneider et al. 2002a; van Uitert et al. 2018)

$$
g_+^l(\theta) = \frac{1}{\theta^2}\left[\theta\ell_{\text{up},l}\,\text{J}_1(\theta\ell_{\text{up},l}) - \theta\ell_{\text{lo},l}\,\text{J}_1(\theta\ell_{\text{lo},l})\right] \; ;
$$
(23)
$$
g_-^l(\theta) = \frac{1}{\theta^2}\left[\mathcal{G}_-(\theta\ell_{\text{up},l}) - \mathcal{G}_-(\theta\ell_{\text{lo},l})\right] \; ,
$$

with

$$
\mathcal{G}_-(x) = \left(x - \frac{8}{x}\right)\text{J}_1(x) - 8\text{J}_2(x) \; .
$$
(24)

While Eq. (20) forms the basis of our estimator, it is convenient for the modelling to link the band powers directly to the angular power spectra via

$$
C_{\text{E},l}^{(ij)} = \frac{1}{2\mathcal{N}_l}\int_0^\infty \text{d}\ell\,\ell\left\{W_{\text{EE}}^l(\ell)\,C_{\epsilon\epsilon,\text{E}}^{(ij)}(\ell) + W_{\text{EB}}^l(\ell)\,C_{\epsilon\epsilon,\text{B}}^{(ij)}(\ell)\right\} \; ; \quad (25)
$$
$$
C_{\text{B},l}^{(ij)} = \frac{1}{2\mathcal{N}_l}\int_0^\infty \text{d}\ell\,\ell\left\{W_{\text{BE}}^l(\ell)\,C_{\epsilon\epsilon,\text{E}}^{(ij)}(\ell) + W_{\text{BB}}^l(\ell)\,C_{\epsilon\epsilon,\text{B}}^{(ij)}(\ell)\right\} \; ,
$$

with kernels given by

$$
W_{\text{EE}}^l(\ell) = W_{\text{BB}}^l(\ell) = \int_0^\infty \text{d}\theta\,\theta\,T(\theta)\left\{\text{J}_0(\ell\theta)\,g_+^l(\theta) + \text{J}_4(\ell\theta)\,g_-^l(\theta)\right\} ;
$$
$$
W_{\text{EB}}^l(\ell) = W_{\text{BE}}^l(\ell) = \int_0^\infty \text{d}\theta\,\theta\,T(\theta)\left\{\text{J}_0(\ell\theta)\,g_+^l(\theta) - \text{J}_4(\ell\theta)\,g_-^l(\theta)\right\} .
$$
(26)

The band power filters in Fourier space (Eq. 26) and in configuration space ($T(\theta)\,g_\pm^l(\theta)$, cf. Eq. 20) are shown in Figs. 2 and 3, respectively, for our default correlation function binning and band power definitions. In particular, we choose eight bands in the range $100 < \ell < 1500$, uniformly spaced in the log. At higher angular frequencies, where many modes are available for a given band, our band powers have a very clean selection in Fourier space, with $W_{\text{EE}}^l(\ell)$ approaching a top hat to very good approximation. Equation (25) shows that band power B-modes could
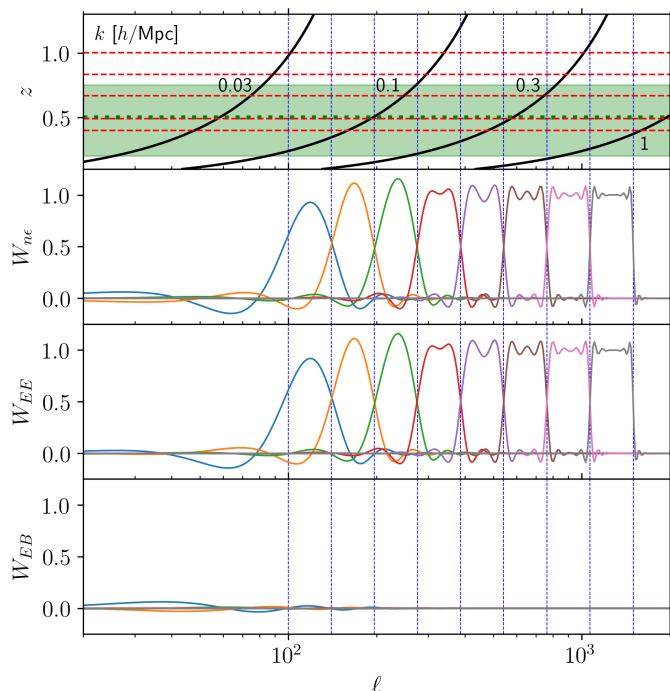

Fig. 2. Bandpower Fourier-space filters for 8 bands logarithmically spaced $\ell \in [100, 1500]$, as indicated by the vertical lines. The second to fourth panels from the top show the galaxy-galaxy lensing, cosmic shear, and EB-mode mixing kernels; see Eqs. (26), and (33). The top panel shows redshift z as a function of ℓ for a selection of wavenumbers k given in units of $h\,\text{Mpc}^{-1}$. Red horizontal lines mark the mean redshifts of the source redshift bins, and the green band shows the redshift range of the two lens samples with a split at $z = 0.5$. The apodisation width is chosen as $\Delta_x = 0.5$.

in principle still be generated even if $C_{\epsilon\epsilon,\text{B}} \equiv 0$, but the mode mixing via $W_{\text{EB}}^l(\ell)$ is negligible.

For the low-ℓ bands these windows become broader while still clearly peaking in the centre of the defined band, with some out-of-band sensitivity especially at $\ell < 100$ (see also Asgari & Schneider 2015 for a detailed analysis of band power filter dependencies). These inherently linear scales are still accurately modelled; the gradual decline in the accuracy of the flat-sky and Limber approximations does not impact our analysis due to the strongly suppressed sensitivity. The corresponding real-space filters shown in Fig. 3 are highly oscillatory and therefore require fine binning of the input correlation functions. For the scales used in our analysis the apodisation only significantly affects large angular scales, reaching 50 % signal suppression at 300 arcmin in this case.

Figure 4 demonstrates that the recovery of the underlying angular power spectra by the band powers is excellent, so that we can use the band powers directly as a faithful representation of the binned angular power spectra in the data. We emphasise however that the translation from true power spectrum to band power according to Eq. (25) and the associated mode mixing is fully taken into account in all quantitative modelling of both the signal and its covariance (in contrast to the earlier work by van Uitert et al. 2018).

Figure 4 also translates the residual uncertainties in matter power spectrum modelling discussed in Sect. 2.1 to band powers. Due to the line-of-sight projection, effects at certain k-scales are spread out over a wider range of angular frequencies, such that all band powers are impacted to some degree. While the limited
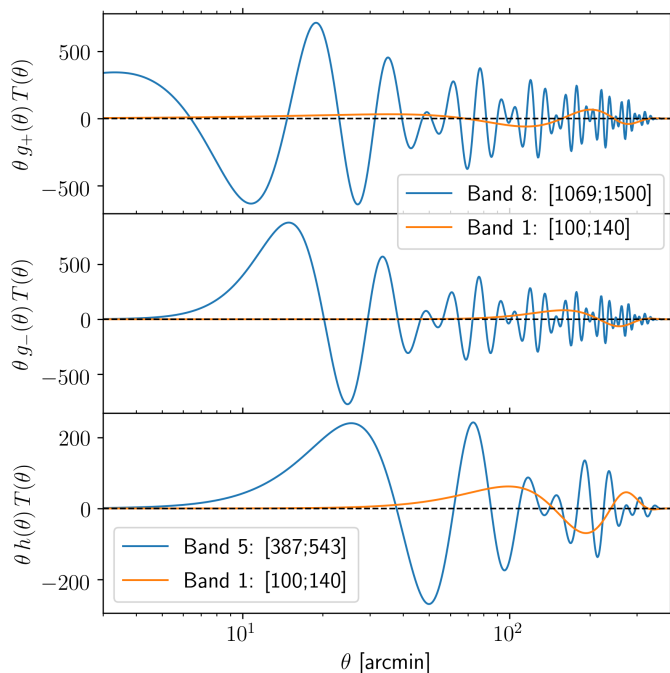
Fig. 3. Bandpower real-space filters for different bands with the angular frequency range given in square brackets. For each case we show the lowest and highest band used in the fiducial analysis setup. The default correlation function binning has been assumed, with apodisation centred on the angular range boundaries of 0.5 arcmin and 300 arcmin, using a log-width of $\Delta_x = 0.5$. *Top*: filter for ξ_+; *centre*: filter for ξ_-; *bottom*: filter for γ_t; cf. Eqs. (22), (23), and (34).
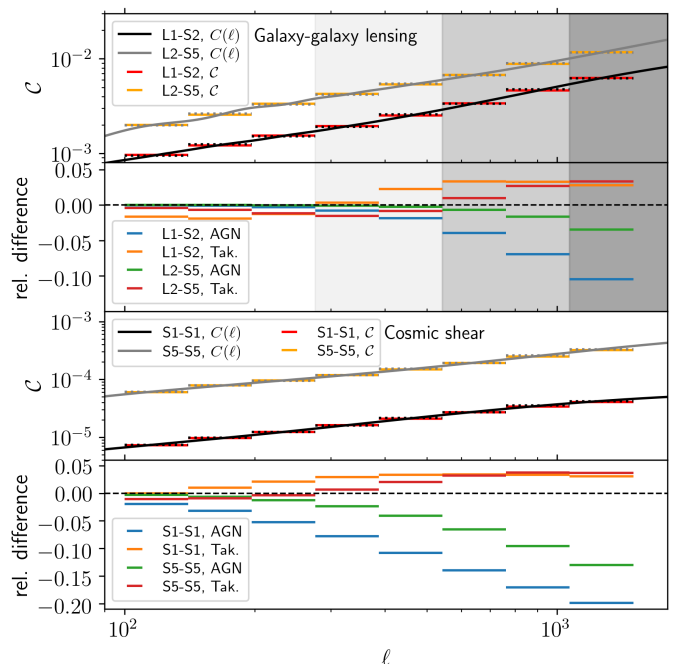
Fig. 4. Bandpowers for galaxy-galaxy lensing (top panels) and cosmic shear (bottom panels) for a selection of redshift bin combinations (indicated by 'L' and 'S' for lens and source galaxy samples, respectively; see Sect. 3 for details). Grey and black lines show the true underlying angular power spectrum; the associated dotted lines the average of these power spectra over the angular frequency band. The bottom panels show the relative deviations in the cases of maximum AGN feedback (blue/green) and using the Takahashi et al. (2012) Halofit version (orange/red) with respect to Mead et al. (2015) without AGN feedback. The grey shaded regions in the top panels show the conservative limits applied to the low- and high-redshift lens samples (excluding all and light/medium grey regions, respectively), as well as the more optimistic cut if galaxy biasing were well understood (only excluding the dark grey region).

accuracy in non-linear power spectrum modelling as indicated by the difference between the Takahashi et al. (2012) and Mead et al. (2015) non-linear prescriptions is at the few per-cent level and close to the limit of tolerable inaccuracy for the current constraining power of data (see the discussion in Sect. 2.1), the potentially large suppression of power by AGN feedback warrants the inclusion of at least one nuisance parameter in the likelihood analysis.

2.5. Galaxy-galaxy lensing summary statistics

We proceed analogously to the cosmic shear case for GGL, writing the angular power spectra as

$$C_{\mathrm{n}\epsilon}^{(ij)}(\ell) = C_{\mathrm{gG}}^{(ij)}(\ell) + C_{\mathrm{gI}}^{(ij)}(\ell) + C_{\mathrm{mG}}^{(ij)}(\ell) , \qquad (27)$$

where i now indexes lens galaxy samples, and j source samples. We include three contributions: the cross-correlation between the lens galaxy distribution and the source gravitational shear ('gG') which contains most of the cosmologically relevant information, the alignment of source galaxies physically close to foreground lenses ('gI'), and the correlation between gravitational shear and the lensing-induced magnification bias in the lens sample ('mG'). The latter term is discussed in detail in Appendix B. We have ignored potential contributions from IA-magnification bias correlations, which would not exceed the per-cent level (Joachimi & Bridle 2010).

The line-of-sight projections read

$$C_{\mathrm{g\,a}}^{(ij)}(\ell) = \int_0^{\chi_{\mathrm{hor}}} \mathrm{d}\chi \, \frac{n_{\mathrm{L}}^{(i)}(\chi) \, W_{\mathrm{a}}^{(j)}(\chi)}{f_{\mathrm{K}}^2(\chi)} \, P_{\mathrm{gm}}\left(\frac{\ell+1/2}{f_{\mathrm{K}}(\chi)}, \chi\right) \qquad (28)$$

for $\mathrm{a} \in \{\mathrm{I, G}\}$, where $n_{\mathrm{L}}^{(i)}(\chi)$ is the comoving distance distribution of lens sample i. The three-dimensional galaxy-matter power spectrum is given by Eq. (3) and the source sample kernels by Eqs. (15) and (16). Since we project along the line of sight over broad line-of-sight galaxy distributions, we do not have to take redshift-space distortions into account. The magnification bias term is modelled as

$$C_{\mathrm{mG}}^{(ij)}(\ell) = 2\left(\alpha_{\mathrm{mag}}^{(i)} - 1\right) C_{\mathrm{GG}}^{(ij)}(\ell) , \qquad (29)$$

where $\alpha_{\mathrm{mag}}^{(i)}$ is usually understood as the power-law slope at the faint end of the luminosity function of a flux-limited sample i (Bartelmann & Schneider 2001). However, the lens samples that we employ have a complex selection function which deviates substantially from a simple flux limit in a single band. As detailed in Appendix B, we have developed and validated an approach to measure an *effective* luminosity function slope α_{mag}, which can be employed in the standard formalism.

Assuming a top-hat band power response as in the cosmic shear case, GGL band powers can be expressed in terms of the tangential shear correlation function,

$$C_{\mathrm{n}\epsilon,l}^{(ij)} = \frac{2\pi}{\mathcal{N}_l} \int_0^\infty \mathrm{d}\theta \, \theta \, \langle\gamma_{\mathrm{t}}\rangle^{(ij)}(\theta) \, T(\theta) \, h^l(\theta) , \qquad (30)$$

with

$$\langle\gamma_{\mathrm{t}}\rangle^{(ij)}(\theta) = \int_0^\infty \frac{\mathrm{d}\ell\,\ell}{2\pi} \, C_{\mathrm{n}\epsilon}^{(ij)}(\ell) \, \mathrm{J}_2(\ell\theta) . \qquad (31)$$

Table 1. Effective survey areas and sky overlap. The overlap is also illustrated in Fig. 5.

	KiDS-1000	BOSS	2dFLenS
survey area	773.3^a	9329.0^b	510.8^c
KiDS overlap	-	319.5	341.9

Notes. All survey areas given in $\deg^2$. Areas involving BOSS are extracted from a HEALPIX map with $N_{\text{side}} = 2048$ (due to the limited size of the random catalogue used to construct the footprint); all other areas are obtained from $N_{\text{side}} = 4096$ maps. The KiDS overlap is limited to full nine-band KiDS+VIKING coverage. *a*) Unmasked area with nine-band coverage. *b*) From Alam et al. (2017). *c*) Using only survey area adjacent to KiDS-S.

Likewise, they can be directly calculated from the angular power spectrum,

$$C_{\text{n}\epsilon,l}^{(ij)} = \frac{1}{\mathcal{N}_l} \int_0^\infty d\ell\,\ell\,W_{\text{n}\epsilon}^l(\ell)\,C_{\text{n}\epsilon}^{(ij)}(\ell)\,, \tag{32}$$

with the kernel

$$W_{\text{n}\epsilon}^l(\ell) = \int_{\theta_{\min}}^{\theta_{\max}} d\theta\,\theta\,T(\theta)\,J_2(\ell\theta)\,h^l(\theta)\,. \tag{33}$$

Here, we have followed van Uitert et al. (2018) in defining the function

$$\begin{aligned} h^l(\theta) = -\frac{1}{\theta^2}\Big[&\theta\ell_{\text{up},l}J_1(\theta\ell_{\text{up},l}) - \theta\ell_{\text{lo},l}J_1(\theta\ell_{\text{lo},l}) \\ &+ 2J_0(\theta\ell_{\text{up},l}) - 2J_0(\theta\ell_{\text{lo},l})\Big]\,. \end{aligned} \tag{34}$$

Figure 2 also displays the $W_{\text{n}\epsilon}^l$ kernel whose form is very similar to the mode-preserving kernel (first equality of Eq. 26) for cosmic shear. The top panel illustrates that wavenumbers above $k = 1\,h\,\text{Mpc}^{-1}$ do not contribute significantly to the band power signals except for Band 8. We therefore study an optimistic scenario that includes Bands 1-7 ($\ell < 1069$) for GGL in the inference. However, by default we restrict the analysis to Bands 1-3 ($\ell < 276$) in the low-redshift lens bin (L1) and Bands 1-5 ($\ell < 543$) in the high-redshift bin (L2), which strongly suppresses all highly non-linear scales beyond $k = 0.3\,h\,\text{Mpc}^{-1}$ (cf. the grey shading in Fig. 4). For completeness the GGL real-space filter is shown in the bottom panel of Fig. 3. The conclusions from Fig. 4 are analogous to the cosmic shear case: excellent recovery of the underlying power spectrum by the band powers and similar impact of residual matter power spectrum uncertainties on the bands.

3. Data and measurements

In this section we summarise the characteristics of the surveys involved and provide an overview of the data processing up to the level of the summary statistics that enter the likelihood analysis. Figure 5 shows the footprints of the KiDS, BOSS, and 2dFLenS surveys. Table 1 provides the corresponding survey areas and the respective overlaps.

3.1. BOSS and 2dFLenS

Our analysis relies on two spectroscopic galaxy surveys, the Sloan Digital Sky Survey (SDSS)-III BOSS (Eisenstein et al. 2011) and the 2dFLenS (Blake et al. 2016) surveys, which cover

Table 2. Galaxy sample properties for BOSS/2dFLenS lens ('L') samples and KiDS-1000 source ('S') samples.

Bin ID	z range	mean z	n_{eff} [arcmin^{-2}]	$\sigma_{\epsilon,i}$
L1	[0.2, 0.5]	0.38/0.36	$14.5/5.8\times10^{-3}$	-
L2	[0.5, 0.75]	0.61/0.60	$16.6/6.1\times10^{-3}$	-
S1	[0.1, 0.3]	0.39	0.85	0.28
S2	[0.3, 0.5]	0.49	1.56	0.27
S3	[0.5, 0.7]	0.67	2.23	0.28
S4	[0.7, 0.9]	0.83	1.52	0.27
S5	[0.9, 1.2]	1.00	1.39	0.28

Notes. Redshift ranges for the source samples are in terms of the best photometric redshift estimate z_B. The first (second) entries for lens mean redshifts and n_{eff} are for BOSS (2dFLenS). The galaxy number density n_{eff} is given in units of arcmin^{-2}; $\sigma_{\epsilon,i} = \sigma_\epsilon/\sqrt{2}$ is the dispersion per ellipticity component. We note that the source sample properties were extracted from KV450 as well as blinded KiDS-1000 data but are expected to be representative of the final KiDS-1000 data. The mean redshift of bin S1 lies outside its z_B range, which is due to an extended high-redshift tail; see Fig. 6.

disjoint areas of sky and were conducted from the northern and southern hemispheres, respectively. The clustering measurements used in our analysis were made over the full BOSS area but do not employ 2dFLenS, whereas the GGL measurements use lenses from both surveys with roughly the same contribution in terms of sky area to maximise the overlap with KiDS.

We use the same galaxy samples as S17 who based their analysis on the final data release of BOSS, DR12 (Alam et al. 2015), which combined the two main BOSS galaxy samples, LOWZ and CMASS, to obtain contiguous redshift coverage between $z = 0.2$ and $z = 0.75$[7]; see Fig. 6, upper panel. As in the direct measurements of baryon acoustic oscillations and redshift-space distortions in the final BOSS data, S17 incorporated an additional $\sim 1000\,\deg^2$ of LOWZ in the Northern Galactic Cap with modified target selection that yielded a lower number density of successful redshift measurements, bringing the effective survey area of both the LOWZ and CMASS samples to ca. $9300\,\deg^2$ (Reid et al. 2016). The LOWZ and CMASS samples are dominated by massive early-type galaxies, but unlike other sample selections of luminous red galaxies (LRGs), they contain a significant fraction of spiral galaxies as well as of order $10\,\%$ satellites (White et al. 2011; Masters et al. 2011).

Corrections for observational systematics in the BOSS samples (Reid et al. 2016; Ross et al. 2017) are applied. Potential residual systematics would primarily affect large scales, but Tröster et al. (2020a) showed that removing these from the analysis does not impact on the cosmological constraints or the goodness of fit. The range over which the clustering model is valid was studied in detail by S17, leading to the choice of $20\,h^{-1}\text{Mpc}$ as a conservative minimum scale. Tröster et al. (2020b) confirmed that this scale remains appropriate when analysing the data in the more constrained ΛCDM framework as opposed to the generic summary parameters employed by S17. The addition of weak lensing data could in principle further tighten the accuracy requirement. However, the mock results in Fig. 1 of Tröster et al. (2020b) show that any systematic trends would shift the posterior orthogonally to S_8 and thus the direction in parameter space impacted by cosmic shear. Hence, we conclude the same clustering scales as in S17 can also be employed in the joint analysis with KiDS-1000.

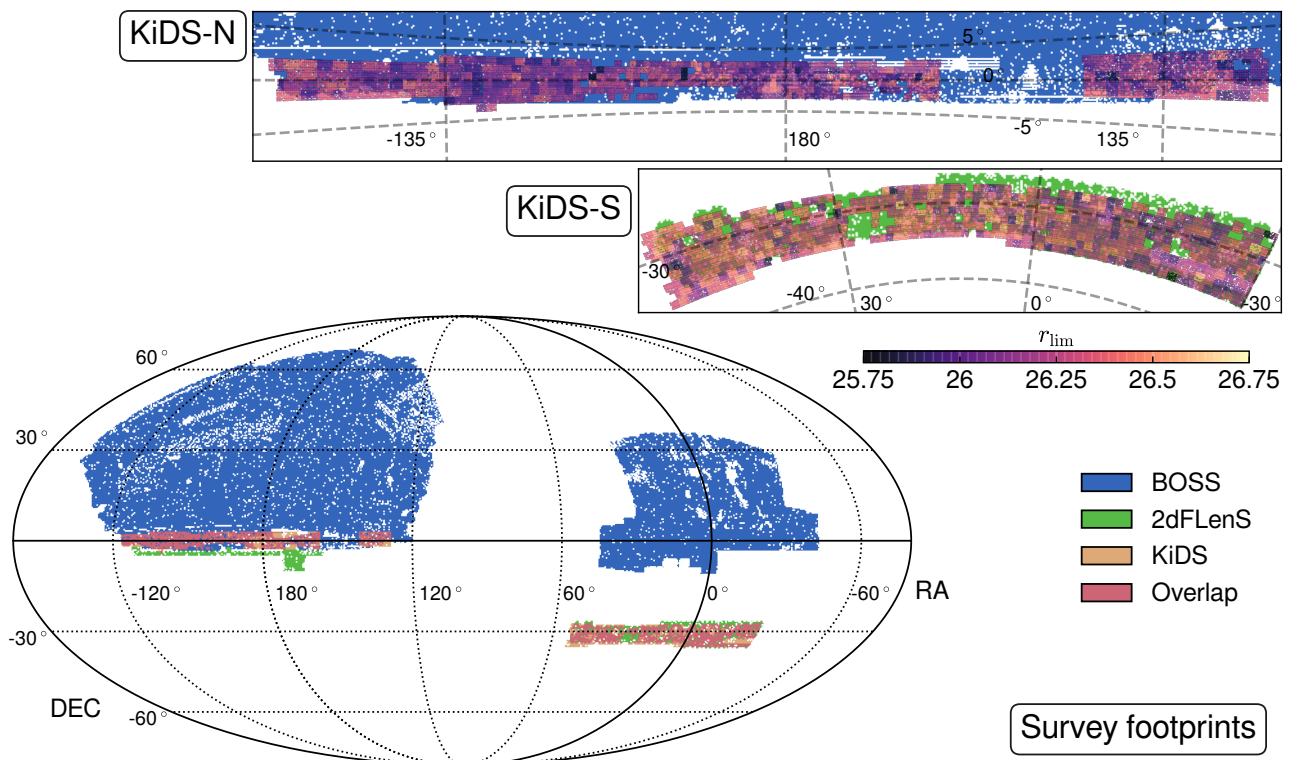[7] Data access: `data.sdss.org/sas/dr12/boss/lss`

Fig. 5. *Bottom*: Survey footprints of three data sets used in this analysis: BOSS (blue), 2dFLenS (green), and KiDS (orange). Overlapping regions of KiDS with either BOSS or 2dFLenS are shown in pink. *Top*: Cut-outs of the KiDS areas showing the variable depth patterns of our source samples as indicated by the r-band magnitude limit, $r_{\rm lim}$ (at 1σ for adaptive aperture size; see Kuijken et al. 2019 for details).

2dFLenS[8] was conducted at the Anglo-Australian Telescope, designed to complement BOSS in southern regions of deep imaging surveys used for weak lensing studies. The survey totals $731\,{\rm deg}^2$ of which we use the largest contiguous portion that overlaps the southern KiDS field. The target selection for its LRG-like samples follows that of BOSS but relies on the VST ATLAS survey (Shanks et al. 2015) for target imaging. The resulting galaxy samples are similar to their BOSS counterparts, but sparsely sampled at $\sim 40\,\%$ of BOSS's galaxy number density (see Table 2 and Fig. 6).

We combine the GGL measurements in the BOSS and 2dFLenS overlaps weighted by the respective galaxy pair counts. This corresponds to inverse variance weighting in the limit that sample variance is negligible. The redshift distributions are combined consistently with that approach (see the green histogram in Fig. 6), and the combined statistics are treated as a single measurement in the subsequent analysis. We refrain from deriving potentially slightly more optimal weights from the full covariance (see Sect. 5) as this would complicate the data processing and introduce a cosmology dependence in these weights through sample variance contributions. We considered re-weighting the 2dFLenS samples to obtain a closer match to the BOSS redshift distribution, but decided not to do so as this would have further reduced the effective number density of 2dFLenS. Re-weighting 2dFLenS was found to have no impact on the likelihood analysis. We define two lens galaxy samples, denoted by L1 and L2, cutting at $z = 0.5$. The third redshift bin in the range $[0.4; 0.6]$ used by S17 is not considered as it is almost fully correlated with L1 and L2.

3.2. KiDS-1000

The Kilo-Degree Survey (KiDS) was conducted from 2011 to 2019 at the 2.6 m VLT Survey Telescope (VST) using Omega-CAM (Kuijken 2011). This analysis makes use of Data Release 4 (DR4; Kuijken et al. 2019), comprising essentially all tiles with *ugri*-band data taken up to January 2018 and covering a raw survey area of $1006\,{\rm deg}^2$. By design, KiDS has the same footprint as the VISTA Kilo-degree INfrared Galaxy survey (VIKING; Edge et al. 2013), together providing deep nine-band photometry in the *ugriZYJHK*$_{\rm s}$ bands. DR4 is a complete re-reduction of the KiDS data set with the VIKING reduction of Wright et al. (2019) fully incorporated, and more than doubles the area with respect to DR3.

Earlier KiDS data releases, which cosmological analyses so far have been based on, still had quite fragmented sky coverage while DR4 is fairly homogeneous in three larger patches, two of which are equatorial with small separation (KiDS-N) and one that straddles Dec $= -30$ deg (KiDS-S); cf. Fig. 5. Correlation function measurements from these patches are combined through galaxy pair weighting and further processed as a single statistic, with redshift distributions, shear calibration, galaxy number densities, etc. averaged accordingly.

KiDS data is processed with two pipelines: Astro-WISE (McFarland et al. 2013) and dedicated software for joint flux measurements in KiDS and VIKING images (see Wright et al. 2019) produce coherent nine-band photometry particularly relevant for photometric redshift estimation (see Sect. 3.3); Theli (Erben et al. 2005) is optimised for weak lensing shear measurement and run on the *r*-band images only. KiDS benefits from a survey and instrument design that is optimised for weak lensing studies, with benign PSF variations across the $1\,{\rm deg}^2$ field of view and a mean *r*-band seeing of 0.7 arcsec in DR4. Improve-

---

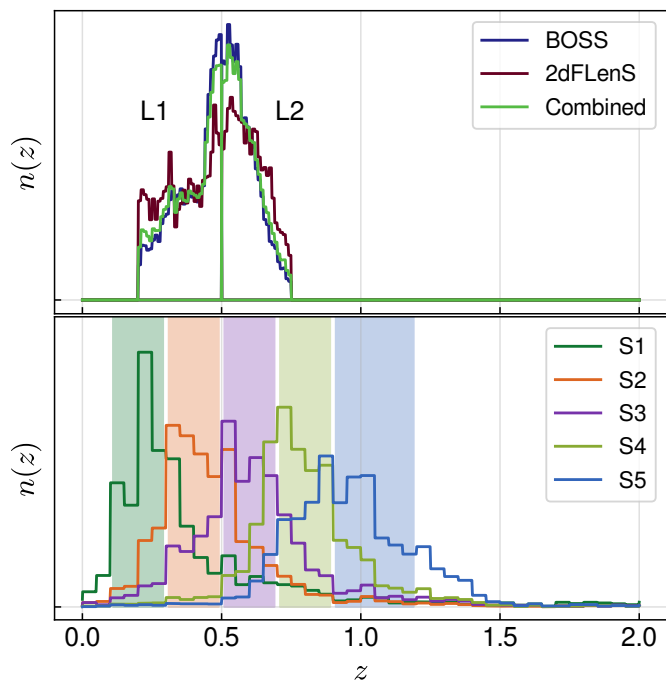[8] Data access: 2dflens.swin.edu.au/data.html

Fig. 6. Redshift distributions of the lens (upper panel) and source (lower panel) galaxy samples. Lens galaxies are split at $z = 0.5$ into two redshift bins for both the BOSS and 2dFLenS surveys. Modelling is carried out with a weighted combination of their redshift distributions (in green). Lens distributions are constructed for galaxies in the KiDS overlap regions, respectively. Source galaxies are split into five tomographic bins based on their best photometric redshift estimates, as indicated by the shaded bands. The source redshift distributions are KiDS-1000-like and built from KV450 data.

ments in the more recent observations mean that DR4 has overall lower PSF variability across the survey than DR3. Gravitational shear estimates are obtained with the model-based *lens*fit (Miller et al. 2007, 2013; Fenech Conti et al. 2017) software, with only minor changes compared to previous KiDS analyses (see KV450). For details of the image processing, nine-band photometry, and shear catalogue production we refer the reader to Kuijken et al. (2019); Wright et al. (2019) and Giblin et al. (2020), respectively.

3.3. Photometric redshifts and tomographic binning

Photometric redshifts are determined from the joint KiDS and VIKING photometry with the Bayesian template-fitting code BPZ (Benítez 2000) as detailed in Wright et al. (2019). The output maximum posterior redshift z_B is used to divide the source galaxy sample into five tomographic bins with boundaries {0.1; 0.3; 0.5; 0.7; 0.9; 1.2}, as in KV450. The photometric redshifts are not used for any other purpose in the KiDS analysis.

The source redshift distributions $n_S(z)$ of the tomographic bins, shown in the lower panel of Fig. 6, are determined from a compilation of spectroscopic redshift data sets that overlap with KiDS-1000 and dedicated calibration observations with the VST in additional fields with deep VISTA and spectroscopic survey coverage (Wright et al. 2019). Totalling more than 25 000 objects, the spectroscopic redshift collection is then re-weighted to become representative of the KiDS source galaxy distribution. In a major update from previous analyses, we adopt the methodology presented in Wright et al. (2020a) applied to DR4, which means that, rather than defining spectroscopic representation via
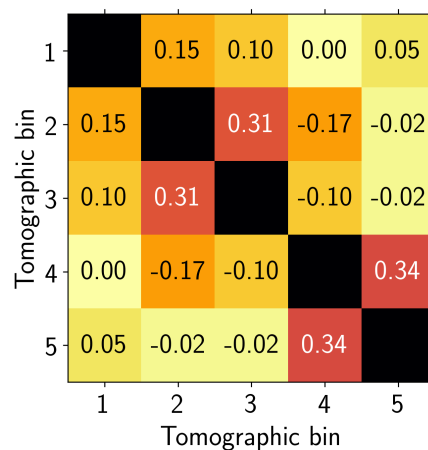
Fig. 7. Correlations between the shifts of the source redshift bins S1–5 arising from the SOM calibration. The covariance was estimated from multiple mock realisations as described in Wright et al. (2020a).

a nearest-neighbour search in the nine-dimensional magnitude space of KiDS+VIKING (referred to as the DIR approach), we now use an unsupervised clustering approach employing a self-organising map (SOM, Kohonen 1982). The SOM creates a two-dimensional representation of colour space onto which both the spectroscopic and source galaxy samples are mapped. The relative abundance of the two samples in a given cell of the SOM space is then used to derive the re-weighting.

The procedure is applied to each tomographic bin individually to reduce the risk of misrepresenting multiple galaxy populations that overlap in colour space by a small number of spectra. Contrary to the previous DIR method, regions without any spectroscopic representation can now be removed from the source sample altogether to create a 'gold' sample with secure calibration. For KV450, this step reduces the effective number density of galaxies by $15 − 17\,\%$ in S1–4 and $6\,\%$ in S5, which however is partially offset by a reduced uncertainty in the centring of the $n(z)$ (cf. Wright et al. 2020b).

As detailed in Wright et al. (2020a), 100 mocks based on quasi-independent lines of sight in the MICE2 simulations (Carretero et al. 2015; Crocce et al. 2015; Fosalba et al. 2015a,b; Hoffmann et al. 2015) are used to assess the fidelity of the SOM approach outlined above. Biases in the mean of the redshift distributions are small, $\delta_z \lesssim 0.01$, with standard deviations of

$$\sigma_z = \{0.005; 0.006; 0.006; 0.004; 0.005\} \tag{35}$$

in S1-5, respectively. Moreover, the simulations reveal substantial correlations between the calibration in each bin, caused by the same spectroscopic redshifts being used in the calibration of multiple tomographic bins. While these correlations are strong in the previous nearest-neighbour approach with correlation coefficients around 0.9 for adjacent bins, they are mild for the SOM algorithm; see Fig. 7.

We account for the residual uncertainty in the tomographic redshift distributions by allowing shifts in the mean redshift of each bin with a multivariate Gaussian prior on the vector of $n(z)$ shifts δ_z with the correlation structure determined by Fig. 7. The standard deviations of the prior are given by the values in Eq. (35) but multiplied by a factor 2 to account for additional uncertainty due to the limitations in the representativeness of our MICE mocks (see Hildebrandt et al. 2020b for details). We note that those mocks contain the smaller survey area of the KV450 analysis, but as the method is insensitive to the source survey

area, we do not expect any impact on our results. The calibration of the redshift distributions and the propagation of the residual uncertainty into the likelihood analysis focus on the mean only because biases in higher moments of the distributions are expected to be subdominant (Amara & Réfrégier 2007). In a separate publication we demonstrate this explicitly for a KiDS cosmic shear analysis by propagating the statistical uncertainty of the full redshift distribution functions (Stölzner et al. 2020).

In the KiDS-1000 analysis the δ_z prior will be centred on the mock-based estimates of the bias (Wright et al. 2020a), but for the demonstration of the methodology in this work we keep the mean of the Gaussian at zero. As the calibrated $n_S(z)$ for KiDS-1000 are not produced until a fairly late stage of the data processing, we adopt in this methodology study the source sample properties as measured from KV450 (see Table 2), again with little change expected in the transition to KiDS-1000. For the redshift probability density functions themselves we adopt a KiDS-1000 distribution from one of the blinded shear catalogues (the $n_S(z)$ vary due to the blinded shear estimation weights; see Appendix C), calibrated with the approach described in Wright et al. (2020a). Further details on the redshift calibration for KiDS-1000, including additional validation via cross-correlations with spectroscopic reference samples, can be found in Hildebrandt et al. (2020b).

As Fig. 6 shows, some source bins have substantial probability of galaxies being located at lower redshift than the lens samples L1 and/or L2. Sources in front of lenses do not generate any of the cosmological or astrophysical signals that we model, except for a very weak magnification bias contribution. To prevent any potential unaccounted systematics entering the data set via this route, we therefore exclude the lens-source bin combinations L1-S1, L2-S1, L2-S2, and L2-S3 from the GGL data vector entering the likelihood. We test for the impact of excluding bin combinations with large overlap between lenses and sources (L1-S2 and L2-S4; see Sect. 7), but include them by default as they contain valuable information about intrinsic alignments.

3.4. Shear calibration

We adopt the shear calibration approach of KV450, which builds on the image simulations presented in Kannawadi et al. (2019). The simulations are generated by sampling from galaxy images emulated from superior Hubble Space Telescope observations in the COSMOS field (Griffith et al. 2012), including realistic distributions of colour, size, ellipticity, and their correlations. While the simulations were designed to accurately reproduce galaxy properties per tomographic redshift bin (rather than only for the overall survey), explicit cross-talk between shear measurement selection bias and photometric redshift estimates is not yet taken into account (see e.g. the discussion in Asgari et al. 2019). This requires multi-band image simulations and will be addressed in forthcoming analyses as it is expected to be a second-order effect. Leading-order effects are captured because the COSMOS galaxies are selected based on their z_B value obtained from KiDS photometry. Troxel et al. (2018b) find the residual biases to be more than an order of magnitude smaller than the response of shear to noisy galaxy images (see Appendix C for details), which we can therefore assume to be negligible since our multiplicative shear bias corrections are small to begin with.

Both additive and multiplicative shear bias are generally significant enough to require calibration and uncertainty propagation into the likelihood analysis. It is demonstrated in Giblin et al. (2020) that biases proportional to PSF ellipticity (α_{PSF}) or residual PSF ellipticity after modelling (β_{PSF}) introduce negligi-

ble bias on cosmology for KiDS-1000. Kannawadi et al. (2019) found the multiplicative biases m for the KV450 analysis setup to be at the per-cent level. The m-correction is applied when computing correlation functions, and propagated accordingly into effective number densities n_{eff}, observed ellipticity dispersions σ_ϵ, and galaxy pair counts, as detailed in Appendix C.

No multiplicative bias corrections are applied to the mocks used in this work to demonstrate the methodology, but we do consider the uncertainty in this correction, which constitutes a significant contribution to the total error budget (see Sect. 5). As in KV450, we model the uncertainty as an additive contribution to the covariance of the two-point statistics, assumed to follow a Gaussian distribution with standard deviation σ_m. The only major update with respect to KV450 that could affect shear calibration significantly is the switch to the SOM-derived gold sample. However, re-running the Kannawadi et al. (2019) analysis with the new sample selection yields negligible changes in the shear bias. We therefore work with the σ_m values for the KV450 setup.

We take a more nuanced approach to extracting σ_m values from the Kannawadi et al. (2019) results (see in particular their Fig. 12) than KV450 who chose a blanket value of $\sigma_m = 0.02$ which comfortably envelopes the various simulation configurations tested. We consider the two most extreme settings versus the fiducial case: the case with correlations between galaxy size and ellipticity switched off (as a conservative floor for this dependence), and the case that uses the global galaxy properties rather than those specific to each tomographic bin (as a conservative floor of accounting for the redshift dependence of galaxy properties). The resulting multiplicative biases vary systematically between these two cases as a function of redshift. In each tomographic bin we choose half of the maximum spread between the three cases as the standard deviation of the calibration uncertainty on the fiducial case, with a floor of 0.01 which corresponds to the statistical noise in the fiducial simulation results. This leads to

$$\sigma_m = \{0.019; 0.020; 0.017; 0.012; 0.010\} \quad (36)$$

for S1-5. As in KV450, we assume that the multiplicative bias corrections are fully correlated between the tomographic bins since the σ_m values are determined by systematic trends that smoothly vary across redshift. Kannawadi et al. (2019) identified additional dependencies on the clustering of source galaxies, stellar number density on the sky, and properties of the input COSMOS catalogue, which could act to de-correlate multiplicative biases, but these contributions remain subdominant. Thus, the covariance contribution due to multiplicative shear bias uncertainty reads

$$\text{Cov}_{mult}\left[C_a^{(ij)}; C_b^{(kl)}\right] = C_a^{(ij)} C_b^{(kl)} \quad (37)$$
$$\times \left\{\sigma_m^{(i)}\sigma_m^{(k)} + \sigma_m^{(i)}\sigma_m^{(l)} + \sigma_m^{(j)}\sigma_m^{(k)} + \sigma_m^{(j)}\sigma_m^{(l)}\right\},$$

where $\sigma_m^{(i)} = 0$ for any index i indicating a lens sample in a GGL measurement. The subscripts a and b could either be E/B for cosmic shear or nϵ for GGL. To arrive at this expression, we have assumed that the expectation of m is zero (after correction), and that $\sigma_m^{(i)} \ll 1$ (see Blake et al. 2020 for a more general equation).

Additive shear bias can be directly estimated from the data with some residual statistical uncertainty. Overall, additive shear bias terms are at the few times 10^{-4} level and corrected individually for the two ellipticity components and for KiDS-N/S[9]

---

[9] The two main KiDS patches are calibrated separately as they differ substantially in sky position as well as Galactic latitude, which leads to differences in their observational and astrophysical characteristics, such as the distribution of seeing and the number density of stars.

before further processing of the data. The residual uncertainty in KV450 on the additive bias measurement was 2×10^{-4} which was propagated into the likelihood analysis via a nuisance parameter with a Gaussian prior with scatter given by this error. The global additive bias only enters ξ_+ and particularly affects large angular scales where the cosmological signal is smallest. Ideal band powers derived from correlation functions available over the entire positive real numbers are insensitive to additive shear contributions as those would only affect the monopole. For the finite range of angular scales used for the band powers in this work, we find the additive bias contribution to be negligibly small, reaching at most $\sim 1\,\%$ on large angular scales in the lowest-redshift tomographic bin combination. We therefore do not use an additive bias nuisance parameter (δc in KV450) in the band power analysis.

We also identified a spatially varying additive bias pattern linked to three problematic CCD chips in the OmegaCAM field of view (see Fig. 2 in KV450), which was measured via stellar PSF ellipticities and then propagated into $\xi_\pm$ with a free overall amplitude that was treated as a nuisance parameter. KV450 showed this nuisance parameter to have no effect on cosmological constraints. Since the amplitude of the pattern has not increased in significance in KiDS-1000 and impacts even less on the band powers, we opt to neither correct for it nor include the related nuisance parameter (A_c in KV450). The pattern could in principle generate a characteristic (weak) signal in the B-modes (Asgari et al. 2019), which however are found to be consistent with zero in Giblin et al. (2020).

While our shear calibration procedure takes the impact of blended galaxy images into account, strongly blended objects are likely to be rejected by the shear measurement algorithm or remain undetected if obscured by a foreground object. Since this rejection is more likely to occur along over-dense lines of sight, weak lensing two-point statistics are biased low as a result (Hartlap et al. 2011). However, for the current generation of surveys this effect only causes biases at the per-cent level over relevant angular scales (Harnois-Déraps et al. 2018; see also Samuroff et al. 2018) and can therefore be neglected.

3.5. Correlation function measurements

All summary statistics in this analysis are based on configuration-space correlation function measurements, which have the major advantage that their expectation values are independent of the survey geometries. We adopt the clustering wedge correlation function measurements (Kazin et al. 2012) from S17 whose salient features we summarise below. The consensus analysis in Alam et al. (2017) showed that the S17 results are in good agreement with other 'full-shape' measurements on the same data, using real-space multipole as well as Fourier-space statistics. The wedge-based analyses generally produced tighter parameter constraints, for example the standard deviation on distance was about a third smaller than for the multipole statistics.

Two-dimensional correlation functions $\xi_{\mathrm{gg}}(s, \mu)$ are measured with a Landy-Szalay estimator (Landy & Szalay 1993) in linear bins of separation s with size $5\,h^{-1}$ Mpc and three equidistant bins in μ, totalling 84 data points per redshift bin. The angle entering μ is measured between the separation vector of a galaxy pair and the line of sight at the midpoint of s. As in related BOSS DR12 analyses, s and μ are calculated for a fiducial, spatially flat ΛCDM model with $\Omega_{\mathrm{m}} = 0.31$. For the Landy-Szalay estimator, we adopt the random catalogues used in S17 with an oversampling factor of 50. The correlation functions include a combi-

nation of weights that minimise variance (Feldman et al. 1994), account for redshift failures and fibre collisions, and correct for spatial patterns induced by a number of observational systematics (Ross et al. 2017).

Weak lensing correlation functions are computed using the public TreeCorr[10] (Jarvis et al. 2004) tree code. Its key tuning parameter is `bin_slop`, which sets the accuracy of placing galaxy pairs into the correct angular separation bin measured relative to the bin size. We optimise `bin_slop` to minimise computation time without measurably impacting on the correlation function. Since we measure finely binned correlation functions with more than a hundred bins per decade of angular separation, fairly large values of `bin_slop` (1.5 for cosmic shear and 1.2 for GGL) can be used. Parts of the KiDS-1000 analysis also employ coarsely binned correlation functions; these are obtained by merging the finely binned measurements.

We use the standard estimator for $\xi_\pm$ (Schneider et al. 2002a); see Appendix C for its explicit form including our choice of weights. GGL is measured through the average tangential shear around lenses, estimated via (Mandelbaum et al. 2005)

$$\widehat{\langle \gamma_t \rangle}(\theta) = \frac{\sum_{ls} w_l w_s \, \epsilon_{\mathrm{t}, l \to s} \, \Delta_{ls}(\theta)}{\sum_{rs} w_r w_s \, \Delta_{rs}(\theta)} \mathcal{N}_{\mathrm{rnd}} - \frac{\sum_{rs} w_r w_s \, \epsilon_{\mathrm{t}, r \to s} \, \Delta_{rs}(\theta)}{\sum_{rs} w_r w_s \, \Delta_{rs}(\theta)},$$
(38)

where the sums with indices l, s, and r run over all elements of the lens, source, and random catalogues, respectively. Here, we have used the bin selector function $\Delta_{ij}(\theta)$ which is unity if the angular separation of a galaxy pair indexed by i and j falls into a bin centred on θ, and zero otherwise. The estimator also accommodates different weights w for the lens, random, and source samples as indicated by their index, and $\epsilon_{\mathrm{t}, l \to s}$ denotes the tangential ellipticity of object s measured with respect to the position of object l. The term $\mathcal{N}_{\mathrm{rnd}} := \sum_r w_r / \sum_l w_l$ reduces to the oversampling factor of the random catalogue with respect to the catalogue of lens galaxies for unit weights. We use 100 times more random points than lens galaxies, which we verify reduces any additional noise components due to the finite number of random points to negligible levels (see Appendix E).

The second term in Eq. (38) removes potential spurious contributions to the signal that do not average out due to survey boundaries and masks. It also suppresses non-Gaussian and additive noise contributions to the covariance of the estimator (Singh et al. 2017). The average tangential ellipticity in the first term is not normalised by the effective number of lens-source pairs, but by the number of pairs between the source and random catalogues. This modification is sometimes referred to as the boost correction and removes any signal suppression due to clustering between lenses and sources in case their redshift distributions overlap. We experiment with different weights in the GGL estimator (see e.g. Shirasaki & Takada 2018; Blake et al. 2020), but find that they do not improve signal-to-noise significantly over the basic form of Eq. (38).

3.6. Band power measurements

Band powers can be derived via linear transformation of the correlation functions. For cosmic shear, our E- and B-mode estimators are given by a discretised version of Eq. (20),

$$\hat{C}_{\mathrm{E/B},l}^{(ij)} = \frac{\pi}{\mathcal{N}_l} \sum_k \Delta\theta_k \, \theta_k \, T(\theta_k) \left\{ \hat{\xi}_+^{(ij)}(\theta_k) \, g_+^l(\theta_k) \pm \hat{\xi}_-^{(ij)}(\theta_k) \, g_-^l(\theta_k) \right\},$$

---

[10] `https://github.com/rmjarvis/TreeCorr`

$$(39)$$

where the sum runs over all angular bins of the correlation function, each with bin width $\Delta\theta_k$. The normalisation $\mathcal{N}_l$, the apodisation $T(\theta_k)$, and the kernels $g_{\pm}$ were defined in Sect. 2.4. These expressions would be unbiased estimators of the angular power spectrum in the limit that the sum asymptotes into an integral over the non-negative real numbers and $T \equiv 1$ (Schneider et al. 2002a; van Uitert et al. 2018). In practice the angular extent over which correlation functions can be measured is limited, which is fully accounted for in the modelling of the band powers. Effects of the discretisation in Eq. (39) are negligible as long as the binning is chosen finely enough to smoothly sample the highly oscillatory kernels (see Fig. 3). The corresponding GGL estimator is

$$\hat{C}_{\mathrm{n}\epsilon,l}^{(ij)} = \frac{2\pi}{\mathcal{N}_l} \sum_k \Delta\theta_k\, \theta_k\, T(\theta_k)\, \widehat{\langle\gamma_{\mathrm{t}}\rangle}^{(ij)}(\theta_k)\, h^l(\theta_k)\,, \qquad (40)$$

derived from Eq. (30). Throughout, the superscripts (ij) indicate redshift bin combinations.

We find that 300 logarithmically spaced bins between $0.5\,\mathrm{arcmin}$ and $300\,\mathrm{arcmin}$ (the default range used in KV450) avoids any discretisation effects. This binning scheme is continued for both smaller and larger angular separation to allow for apodisation. An apodisation width of $\Delta_x = 0.5$ (cf. Eq. 22) is found to suppress sensitivity of the band powers far from their designated band width while only mildly expanding the angular range entering the estimators. Consequently, correlation functions in the range $[0.6; 234]$ arcmin enter at full sensitivity, contributions reduce to half at the nominal bin edges of $0.5/300$ arcmin, and scales outside the interval $[0.4; 385]$ arcmin do not contribute at all. The minimum scale is still well above the regime where galaxy blending and the image cutout size with which *lens*fit works ($\lesssim 10\,\mathrm{arcsec}$) could impact on the correlation functions. Conversely, the maximum scale is of the same order as the extent in declination of the KiDS patches (cf. Fig. 5) beyond which galaxy pair counts would rapidly decline and isotropy-breaking systematics in the shear measurement not fully average out.

We choose 8 logarithmically spaced bands with a lower boundary of $\ell = 100$ and an upper limit of $\ell = 1500$. The band power response to lower angular frequencies than $\ell \sim 100$ becomes very broad (see Fig. 2) and would thus necessitate more expensive modelling of very large-scale modes. On scales smaller than $\ell = 1500$ the modelling uncertainty of non-linear matter power and baryonic feedback prevents us from exploiting measurements efficiently. All 8 bands are used in the cosmic shear signals. Our conservative approach to modelling GGL means we only use the lowest three bands for correlations involving L1 and the lowest five bands for those involving L2 (see the discussion in Sect. 2.2). We also explore a more progressive setting that only discards the highest ℓ-band for GGL and thus scales of $k \approx 1\,h\,\mathrm{Mpc}^{-1}$ and smaller.

4. Simulations

We create suites of dedicated simulations with the primary purpose of assessing the accuracy of the two-point statistics we measure and of their covariance. A key question we address is whether the previously used analytic covariance prescription is still good enough for the increasingly constraining data sets of the current survey generation, despite its necessarily idealising assumptions about survey geometry and survey homogeneity.

Non-Gaussian contributions to the covariance are subdominant and have been assessed before by comparison with a suite of N-body simulations (Hildebrandt et al. 2017; Harnois-Déraps et al. 2018), so that we now opt for a fast approximate simulation approach that can accommodate the full survey volume and allows for a large number of mock realisations of the data. The resulting simulation suite has the additional benefit of providing us with accurate sampling distributions of various summary statistics that we generate from the data (see Sect. 6).

4.1. Scope

A number of challenges related to survey characteristics arise in the joint galaxy clustering and weak lensing analysis that necessitate careful validation of signals and covariances with mocks. First, lens and source galaxies occupy vastly different footprints in the sky (see Fig. 5 and Table 1). This complicates the modelling of cross-variances in particular, and the strong assumptions used in analytic modelling to date (cf. van Uitert et al. 2018) are yet to be tested. Secondly, while our choice of two-point statistics are by design insensitive to the survey geometry, it does impact on their covariance. The KiDS footprint with its two disjoint and elongated fields, whose extent in declination is of the same order as the largest scales we measure, could violate the sweeping approximations currently made in the most widely used covariance models.

It is also unclear from first principles how to measure the effective area of a survey that to good approximation scales the variances of all data. One could obtain an accurate estimate of all unmasked area or instead focus more on the extent of the footprint, that is, exclude features like the numerous star masks which are on scales much smaller than those that are considered in the likelihood analysis. This then becomes a question of adequate spatial resolution in the survey mask (either of a pixelated map or a random catalogue). Different choices here can lead to tens of per cent differences in variances and thus outweigh the impact of other contributions like non-Gaussian terms which have received much attention in the literature.

Finally, virtually all analytic modelling has hitherto relied on the assumption of homogeneity, thereby ignoring the spatial variability in the data induced by observing conditions (see Heydenreich et al. 2020, though). However, the depth of an imaging survey varies from pointing to pointing due to changes in conditions such as seeing and sky transparency. Larger-scale patterns can be generated through a combination of seasonal weather patterns and visibility restrictions of targets. The resulting variable survey depths will be more pronounced in surveys like KiDS that visit sky areas only once during the lifetime of the programme. The impact is visualised in the lower part of Fig. 5 where colours ranging from yellow to purple indicate one magnitude of per-pointing variation in the r-band limiting magnitude r_{lim} (1σ detection in adaptive apertures), which is sensitive to both the seeing and the background noise level in the images.

In order to evaluate the impact of multiple survey footprints, mask geometry, and variable depth on our analysis, we require simulations that allow us to impose the realistic footprints and overlaps of the surveys, as well as galaxy samples with precisely the same observational characteristics as in the data. Validation is greatly aided by having a-priori knowledge of the input two-point statistic that is to be recovered, as well as by generating a large number of realisations of the Universe to beat down noise in the estimates of large covariance matrices. These conditions lead us to make use of fast, full-sky random field simulations as described in the following.

We opt for lognormal, rather than Gaussian, random fields because the latter are unable to recover the input two-point statistics at the level of accuracy that we require. At the spatial resolution of our mocks, the KiDS-1000 area is covered by of order 5×10^6 pixels which define volume elements in shells of the matter distribution. If the probability distribution of the matter density contrast δ in these volume elements is Gaussian, values of δ are not bounded from below. Hence, unphysical values of $\delta < -1$ in the tail of the distribution are realised in a small fraction of the pixels. Any modification of these problematic cases, for example by setting them to an empty pixel ($\delta = -1$), leads to significant biases in the two-point statistic of the matter density field at the few per cent accuracy that we aim for. Lognormal statistics of the matter density contrast have the additional benefit of being close to the realistic case (Kayo et al. 2001), which enables us to at least qualitatively explore the non-Gaussian properties in the covariance (see Sect. 5.3).

We emphasise that, by design, our mock suite returns the two-point statistics that we put in. It is therefore not suitable to validate our baseline models of these statistics. This validation was performed in previous works using a range of N-body and hydro-dynamical simulations; see the discussion and references in Sect. 2. In our mocks we do not couple the shear estimates to the foreground galaxy population, so that we do not incorporate neighbour-exclusion bias, which is expected to vary with the local survey depth. However, this effect was shown to be insignificant for the current generation of surveys (Harnois-Déraps et al. 2018). We note that the impact of variable depth on the shear calibration is taken into account in the data analysis; see the discussion in Appendix C.

4.2. Construction of mocks

The matter distribution of our fast mocks is generated with FLASK[11] (Xavier et al. 2016), building concentric shells of lognormal random fields with known angular matter power spectra and a minimum density contrast of -1 (corresponding to an empty line of sight). We note that the same approach to validating covariance models was also taken by Krause et al. (2017). The input cosmology for the mocks is specified in Table A.1. The matter shells are created in 18 redshift bins of roughly equal intervals in comoving distance between $150 - 200 \, h^{-1}$ Mpc. The bin edges closest to the redshift limits of the lens samples are moved to exactly 0.2, 0.5, and 0.75. The correlations between matter shells are fully included via their cross-power spectra. Since the matter slices are fairly thin, the angular matter power spectra are computed using the full curved-sky, non-Limber expressions. Weak lensing shear and convergence fields are then obtained from integrating over the matter shells along the line-of-sight with the weighting of Eq. (15).

We create a new software, SALMO[12], to populate the resulting matter distribution with galaxies based on the visibility masks and redshift distributions provided by the surveys. In contrast to the standard FLASK functionality, SALMO allows us to implement pointing-dependent selection functions and thus spatially varying redshift distributions, galaxy densities, and ellipticity dispersions. Galaxies are sampled following a Poisson process with constant galaxy bias corresponding to the fiducial values of b_1 in the lens samples and then assigned weak lensing signals and shape noise. Tracers with different angular and radial

[11] http://www.astro.iag.usp.br/~flask
[12] Speedy Acquisition of Lensing and Matter Observables; the code will be made public on acceptance of the KiDS-1000 analysis papers.
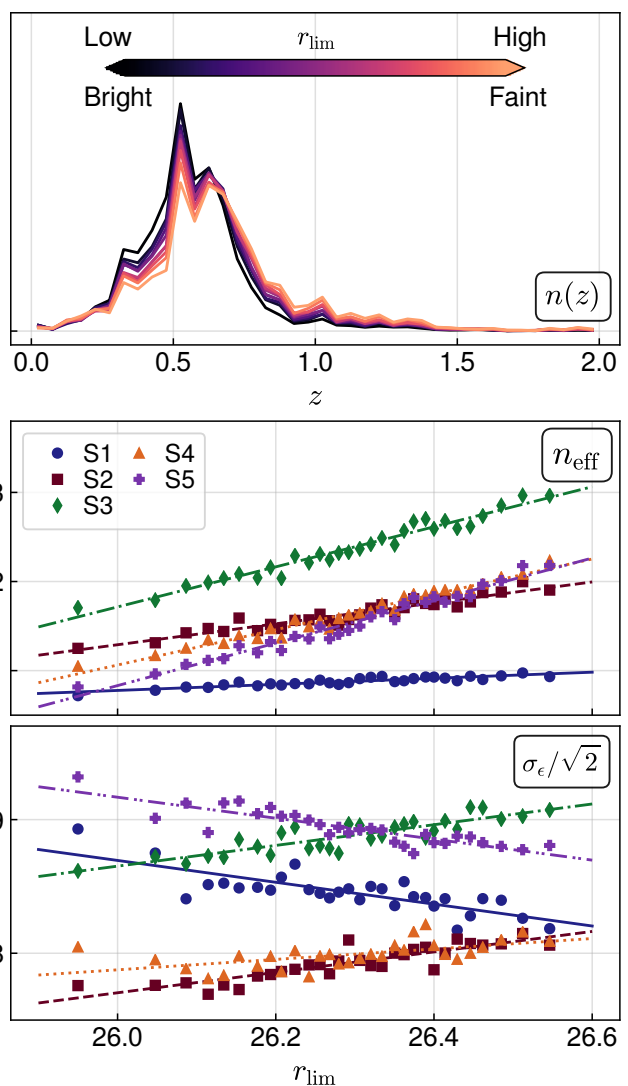
Fig. 8. Characteristic of variable depth effects measured from early KiDS-1000 data. *Top panel:* Redshift distributions of source galaxies in the third tomographic bin (S3) in 10 equi-populated bins of r_{lim} (1σ adaptive aperture size). *Centre panel:* Effective galaxy number density as a function of r_{lim} in the five source galaxy bins, together with linear fits. *Lower panel:* Same as centre panel but for the observed ellipticity dispersion σ_ϵ.

selection functions can thus be simulated based on the same realisation of the underlying matter distribution. Masks for the source galaxy samples are taken from KiDS-1000; those for the lens samples are made from areas of BOSS and 2dFLenS that overlap KiDS (with the exception of the study presented in Appendix D.1, which is based on a set of simulations with the full BOSS footprint). Both masks and the FLASK-generated maps have a HEALPIX[13] (Górski et al. 2005) resolution of $N_{\mathrm{side}} = 4096$, which corresponds to a characteristic angular size of 0.86 arcmin below which the power of density fluctuations realised in the simulations is suppressed.

Lens redshift distributions are either adopted individually from the BOSS and 2dFLenS data, or an inverse variance-weighted combination is applied uniformly to all lenses (cf. Fig. 6, top panel). Since the final KiDS-1000 redshift solutions were still under construction when the mocks were created, we

[13] http://healpix.sourceforge.net

proceed as follows, making use of the default KV450 DIR approach (using a weighting of spectroscopically observed galaxies in the KiDS-VIKING colour space). We divide the $1\,\mathrm{deg}^2$ pointings in KV450 into 10 equi-populated bins of r_{lim} and repeat the redshift distribution calibration on each subset for all five tomographic bins with the full set of spectroscopic sources. The resulting distributions are then assigned to each of the pointings within the respective bins and the mock galaxies are sampled from them accordingly. To limit the size of simulations, we truncate all redshift distributions at $z = 2$ beyond which only $0.3\,\%$ of source galaxies are located. The average of these individual redshift distributions is shown in Fig. 6 and used in the modelling of the data vector.

Figure 8 illustrates the characteristics of variable depth effects for KiDS, as measured from an early version of KiDS-1000 data with blinded ellipticity measurements (which will affect σ_ϵ only at the per-cent level). The upper panel shows the redshift distributions for the different r_{lim} bins in the third tomographic bin (S3). We can see that, as expected, for faint r_{lim} the distribution is shifted towards higher redshifts as better seeing and transparency allow for more faint galaxies to enter the sample, which tend to be at higher redshifts. The number of low-redshift galaxies is only weakly affected, but after normalisation their proportion effectively decreases.

The centre and lower panels of Fig. 8 show the dependence of the effective number density n_{eff} and the observed ellipticity dispersion σ_ϵ on r_{lim}. In this case galaxies from all five tomographic bins are divided into 30 equi-populated r_{lim} bins selected from the KiDS-1000 footprint. We observe a very clear linear dependence for both quantities in all tomographic bins. The galaxy number density increases with deeper data as expected, while there is no clear trend for σ_ϵ, as the sign of the slope reverses for the lowest and highest tomographic redshift bins. We surmise that there is a complex interplay between faint-end galaxies altering the galaxy population and hence the intrinsic ellipticity distribution on the one hand and those galaxies increasing the image level noise which broadens σ_ϵ on the other (one of the reasons why multiplicative shear bias is calibrated per tomographic bin in KiDS; cf. Kannawadi et al. 2019). Based on the linear fits to these trends, we incorporate a per-pointing variation for n_{eff} and σ_ϵ into the mocks as well.

With the methodology above, we produce four sets of mocks[14], following identical underlying dark matter field realisations but differing in their level of realism regarding survey footprints and spatial variability in the redshift distributions: **Buceros** – a single rectangular mask with fully overlapping surveys, spatially uniform lens redshift distributions, and spatially uniform source redshift distributions; **Cygnus** – realistic survey masks, spatially uniform lens redshift distributions, and spatially uniform source redshift distributions; **Diomedea** – realistic survey masks, different lens redshift distributions in KiDS-N/S from BOSS/2dFLenS, and spatially uniform source redshift distributions; and **Egretta** – realistic survey masks, different lens redshift distributions in KiDS-N/S, and variable depth source redshift distributions and galaxy sample properties.

Contrary to full N-body or other particle- and/or mesh-based runs, lognormal random field simulations are extremely fast. Using the setting described above, a set of full-sky tomographic density and lensing maps can be generated within less than an hour of wall-clock time using four CPUs in parallel. Catalogue generation with two BOSS-like lens bins and five KiDS-like

source bins requires about 20 minutes with one CPU. Parallel computation of many survey realisations is then trivial to execute on distributed high-performance computing systems. The speed of the simulation generation shifts the computational bottleneck to the processing of the mock data, in our case driven by the calculation of correlation functions. We create 5000 realisations of each survey setup which we find to reduce the sampling noise in covariance estimates to negligible levels for our comparison.

The fidelity of our mocks is assessed by applying the measurement pipeline to the mock catalogues in the case of uniform galaxy distributions and comparing the resulting two-point statistics to the theoretical predictions. The mock mean agrees to $3\,\%$ or better for both cosmic shear and GGL signals over relevant angular scales, with the exception of cosmic shear correlation functions involving bin S1 for which we see an excess signal around $5\,\%$ over the theoretical model. The latter is due to the small number of redshift shells in our Flask simulations which has the strongest impact on the low-redshift source sample with the smallest redshift baseline in the line-of-sight integration to obtain the lensing signal. This limitation would be easy to remedy, but as the runtime of mocks scales approximately quadratically with the number of shells, we prioritise the ability to create a larger number of mock realisations. While the small excess power in S1 should be borne in mind in the comparison with analytic covariances (see Sect. 5.3), it is of little consequence for the likelihood analysis since the cosmic shear signals including S1 have low signal-to-noise.

4.3. Impact of spatial variability

We run our measurement pipeline on the four variants of mock catalogues and study the relative difference between the resulting band power signals in Fig. 9 (differences in their covariance will be discussed in Sect. 5.3). To suppress noise, we have averaged signals over 1000 realisations and switched off shape noise. Switching from a simple to the realistic survey geometry (Buceros to Cygnus) has no impact on correlation functions and their derived statistics – a major advantage of our choice of harmonic space measure.

Comparing Diomedea to Cygnus shows the effect of averaging GGL over the BOSS and 2dFLenS overlap and modelling the resulting signal with the averaged redshift distribution. This approximation is accurate to within $2\,\%$ and thus fully sufficient for our analysis. We caution however that this conclusion is to a large degree due to the factor 2.5 lower number density in 2dFLenS, which means that the contribution to both the signal and the shape of the redshift distributions is dominated by BOSS (cf. Fig. 6). Averaging over spatial variability in the lens samples, as ubiquitously done when calculating summary statistics, has the potential to cause significant biases in GGL if the parts of the survey(s) that vary in depth carry a more evenly balanced statistical weight than in our case.

The Egretta case additionally incorporates variable depth in the source galaxy samples. Its impact on GGL is small and even partially reverses the small trends due to variability in the lens samples in correlations with L1. The effect on cosmic shear is more pronounced, especially for correlations involving the low-redshift S1 for which deviations of up to $4\,\%$ can be observed at the highest angular frequencies considered. The cosmic shear case was recently studied in depth by Heydenreich et al. (2020) who developed a semi-analytic model and demonstrated very good recovery of Flask/Salmo-generated mock signals. They concluded that on small scales the trend is well accounted for by the pointing-to-pointing variation of survey depth, whereas

---

[14] We choose to name the different mock setups after birds in order to easily distinguish between them.
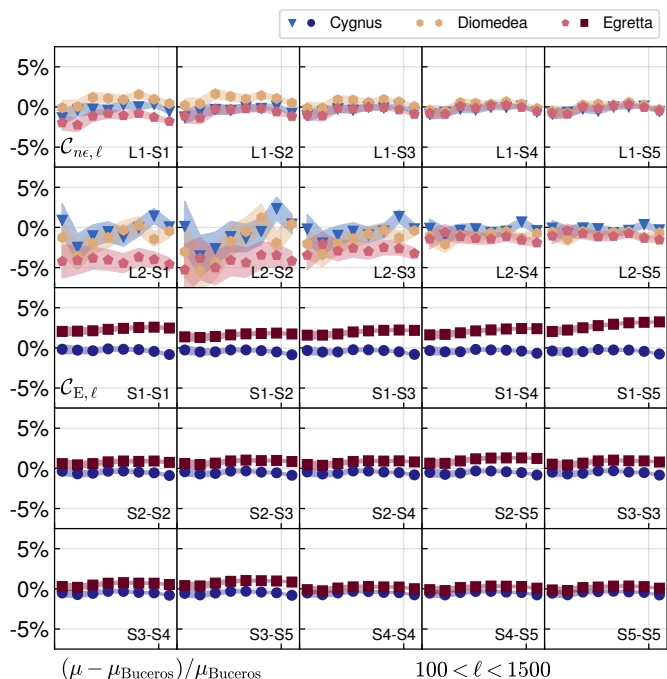
Fig. 9. Comparison of mean estimates of band power between different mock setups. Plotted is the relative deviation from the Buceros case (rectangular mask, uniform depth). Blue, yellow, and red colours correspond to the Cygnus (realistic mask, uniform depth), Diomedea (like Cygnus but using individual lens sample properties for BOSS and 2dFLenS, hence only shown for GGL), and Egretta (realistic mask and depth variations) cases, respectively. The top two rows show GGL signals; the bottom three the cosmic shear signals, with bin combinations indicated in the panels (those not used in the likelihood analysis are greyed out). Bands around the data points indicate the standard error determined from a jackknife estimate of variance on the ratios shown, based on mean signals calculated from 1000 realisations.

above degree scales correlations between the depths of adjacent pointings contribute as well.

The mean redshifts of S1 vary the most, ranging between 0.35 and 0.65 (see Fig. 2 of Heydenreich et al. 2020), leading to the most pronounced variable depth effect. This is likely driven by the extended high-redshift tail of the S1 redshift distribution (Fig. 6) which is strongly affected by the limiting magnitude and drives up the S1 mean redshift well beyond the photometric redshift boundary.

The cosmic shear band powers are consistently enhanced by the variable depth effect. As mean redshifts vary roughly linearly with depth, the amplitude of two-point statistics acquires a quadratic dependence to first order via the lensing efficiencies (cf. Eq. 15), so will tend towards higher amplitudes on average[15]. The source redshift distributions enter the GGL signals linearly, so as long as most of the sources are located behind the lens galaxies, the contributions from the different lines of sight average to the signal based on the mean redshift distribution used for Buceros. However, if the sources are mostly positioned in front of the lenses, it is the high-redshift tails of $n_{\mathrm{S}}^{(i)}(z)$ that determine the signal amplitude, and these have a non-linear dependence on depth (cf. Fig. 8, top panel). This leads to the $\sim 5\,\%$ negative trend in the L2-S1 and L2-S2 GGL signals, which however are not used further in our analysis.

---

[15] Variable depth also generates B-mode cosmic shear signals (Vale et al. 2004), but they are unlikely to be detectable with KiDS.

Since the high signal-to-noise cosmic shear signals involving S3-5 are only affected at the per-cent level, we can expect a mild impact on cosmological parameter constraints overall, with S_8 somewhat over-estimated. Parameters that are sensitive to the small modifications of redshift scaling and angular scale dependence, like A_{IA} and A_{bary}, could in principle be more strongly affected. We add the relative difference between the means of the Egretta and Buceros mocks seen in Fig. 9 to a noiseless mock data vector and carry out a likelihood analysis of the joint weak lensing signals (see Sect. 7), which results in insignificant changes to all model parameters, for instance a 0.12σ shift in S_8.

5. Error modelling

We develop separate covariance models for galaxy clustering and the weak lensing signals, as detailed in the following. This is possible because these parts of the overall data vector are uncorrelated to very good approximation, primarily due to the much larger survey area of BOSS outside the KiDS footprint; see Appendix D.1 and in particular Fig. D.1 for a detailed analysis.

5.1. Clustering covariance

Since we can treat the clustering signal as fully independent of the weak lensing signals, we simply adopt the public S17 covariance that was estimated from 2045 Multidark-Patchy mocks of BOSS (Kitaura et al. 2016). Bias in the mean of the noisy inverse of the covariance estimate entering the likelihood is corrected (Hartlap et al. 2007). S17 also rescaled the posterior to account for noise bias in the second moments of the inverse covariance distribution (Percival et al. 2014), which we do not adopt because it only amounts to a 1.6 % change in parameter uncertainties which is within the numerical uncertainty of posterior sampling. Moreover, this ad-hoc modification would introduce inconsistencies in the constraints jointly with weak lensing as the latter do not suffer noise biases due to using a fully analytic covariance.

5.2. Analytic weak lensing covariance

Analytic approaches to the weak lensing covariance have now been widely adopted by the community as they provide sufficiently accurate, computationally fast, and noiseless models (Abbott et al. 2016; Hildebrandt et al. 2017; Hikage et al. 2019). We adopt the model used for cosmic shear in KV450 with minor updates and extend its use to GGL. A detailed description of the implementation is provided in Appendix E; here, we summarise the salient points.

The model is composed of five terms of cosmological origin,

$$\mathrm{Cov} = \mathrm{Cov}_{\mathrm{G,sva}} + \mathrm{Cov}_{\mathrm{G,mix}} + \mathrm{Cov}_{\mathrm{G,sn}} + \mathrm{Cov}_{\mathrm{NG}} + \mathrm{Cov}_{\mathrm{SSC}} , \quad (41)$$

where 'G' stands for the Gaussian contributions, that is terms that would also be present if summary statistics of a Gaussian random field were considered. The terms split further into a pure noise component ('sn'; covering shot noise due to the finite sampling of the matter density distribution by galaxies and shape noise due to the randomly oriented intrinsic ellipticities of galaxies), a pure sampling variance contribution due to observing a finite volume of the Universe ('sva'), and a component that mixes the two ('mix'). Additional terms arise due to the non-Gaussianity of the underlying fields, which generate higher-order correlations between modes measured by the survey (cap-
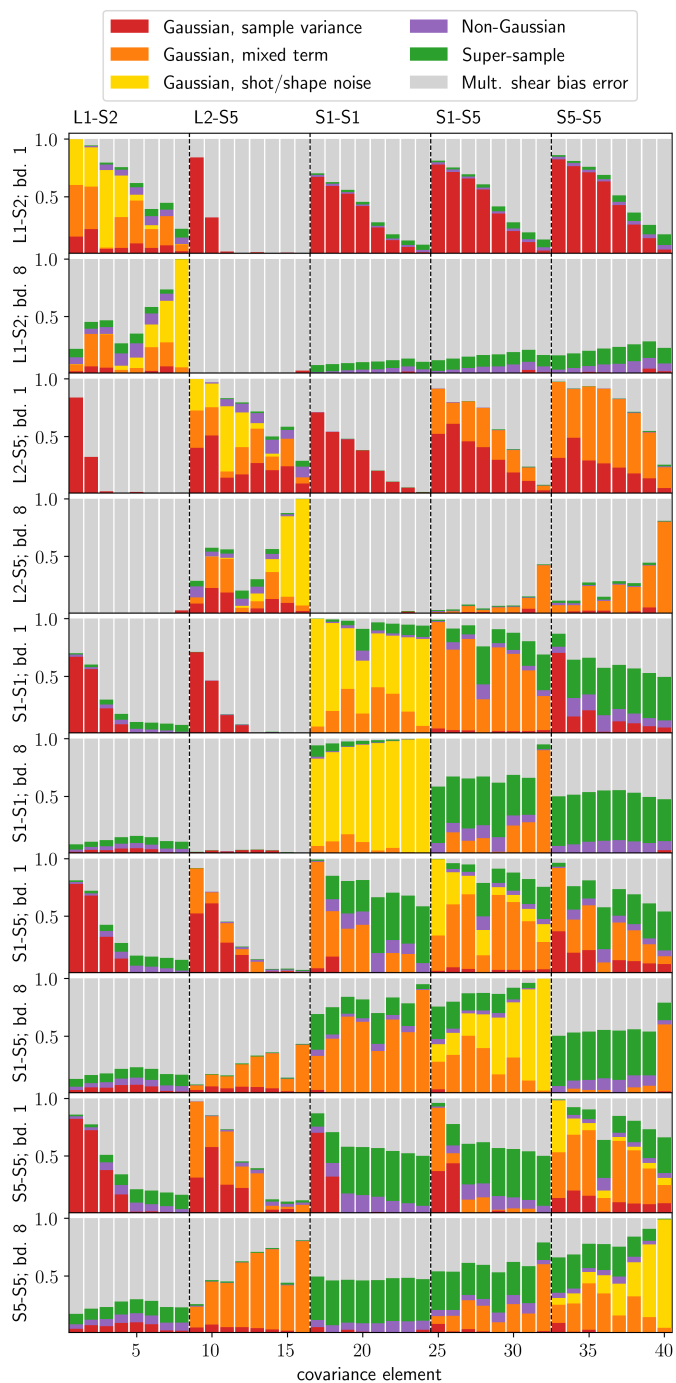
Fig. 10. Contributions to selected elements of the analytic band power covariance. Shown are the relative contributions of the absolute values of the six covariance terms: sample variance (red), mixed term (orange), and shot/shape noise (yellow) in the Gaussian covariance; non-Gaussian in-survey covariance (purple) and super-sample covariance (green); and the multiplicative shear bias uncertainty (grey). Each panel corresponds to a single row in the covariance matrix of the full GGL and cosmic variance data vector, for the signal and angular frequency band indicated in the label; columns in each panel have the same ordering of signals but show all 8 bands, respectively. A representative subset of signals was chosen: two GGL signals, as well as a low- and high-redshift cosmic shear signal, and the corresponding redshift bin cross-correlation.
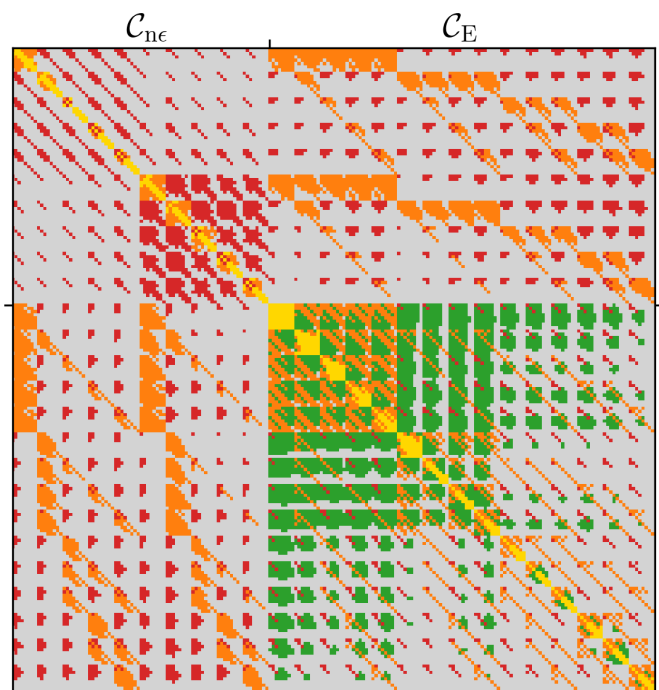
Fig. 11. Dominant contribution (absolute value) to the band power covariance. The colour coding is identical to Fig. 10, i.e. sample variance (red), mixed term (orange), and shot/shape noise (yellow) in the Gaussian covariance; non-Gaussian in-survey covariance (purple, always sub-dominant) and super-sample covariance (green); and the multiplicative shear bias uncertainty (grey). The top left block contains the covariance of the GGL signals ($\mathcal{C}_{n\epsilon}$; without cuts in tomographic bins or angular scales), the bottom right block the cosmic shear covariance ($\mathcal{C}_E$). Smaller blocks discernible correspond to the different bin combinations, and within these, individual 8×8 blocks correspond to the eight band powers of the signal for a single tomographic bin combination.

contribution accounting for statistical uncertainty due to the multiplicative shear bias calibration (see Eq. 37) completes our analytic description.

We calculate real-space correlation function covariances which are subsequently transformed to band powers analogously to the two-point statistics measurements (see Appendix E.3 for the explicit formulae). The modelling of the Gaussian terms follows the approach of Joachimi et al. (2008), with the exception of the noise term which takes the actual galaxy pair counts measured from the data (cf. Appendix E.1) as opposed to working with the expectation for a uniform survey (cf. Troxel et al. 2018a). The non-Gaussian components are constructed through a halo model in line with earlier approaches (Takada & Hu 2013; Li et al. 2014; see also Krause & Eifler 2017). We refrain from implementing full halo occupation statistics for the galaxy clustering covariance terms entering the GGL covariance and its cross-correlations, and only include linear bias via b_1. Higher-order bias terms modify the power spectra by less than 10%, and the strongest deviations occur on scales where noise terms increasingly mitigate inaccuracies in the sample variance terms.

Figure 10 illustrates the contributions of the six model components to a representative subset of covariance elements. Since absolute values are plotted, the components may act to partially cancel each other. Diagonal elements are generally dominated by noise, especially on small scales. Only the GGL variances at low ℓ have an almost equal contribution from noise, Gaussian sample variance, and the mixed term. Away from the diagonal, the

tured in the 'NG' term), as well as between modes within the survey and those on scales larger than the survey footprint (known as super-sample covariance; 'SSC'). A sixth, non-cosmological
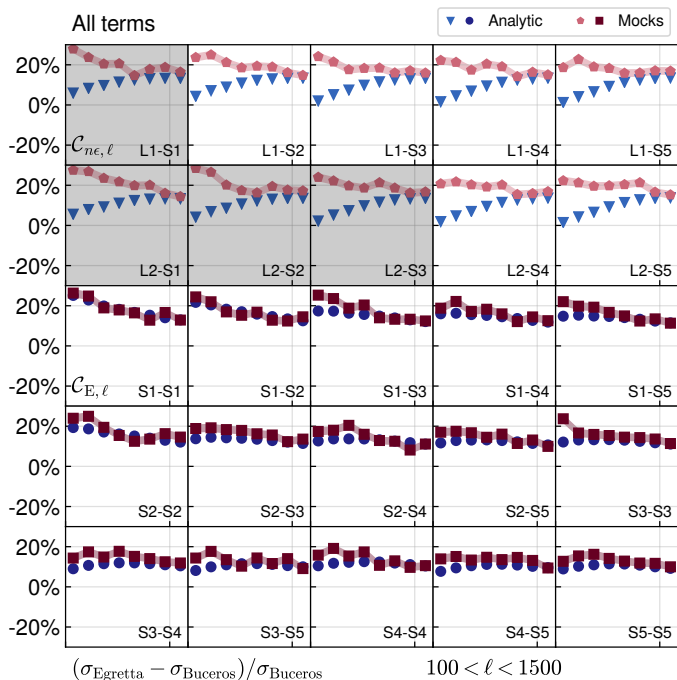
Fig. 12. Relative difference between the standard deviations of different mock setups for the covariance of the weak lensing band power signals. The relative deviation of the Egretta case (realistic footprint and depth variations) from the Buceros case (rectangular mask, uniform depth) is displayed. Red (blue) symbols show results for the mock (analytic) covariance. The top two rows show GGL signals; the bottom three cosmic shear signals, with bin combinations indicated in the panels. GGL signals that are not used in the analysis have been greyed out. Bands around the mock data points indicate the standard error determined from a jackknife estimate of variance.

Fig. 13. Relative difference between the square root of the diagonals of the mock and analytic covariances for the weak lensing band power signals. The top two rows show GGL signals; the bottom three cosmic shear signals, with bin combinations indicated in the panels. Three cases are shown: spatially uniform galaxy samples in a simple footprint (Buceros, green); spatially uniform galaxy samples in the realistic survey footprints (Cygnus, blue); and spatially varying depth in the realistic footprints (Egretta, red). GGL signals that are not used in the analysis have been greyed out.

mixed term and super-sample covariance have strong impact on the level of correlation for cosmic shear. The multiplicative bias term can also reach contributions at the 50 % level, while for cross-correlations involving GGL signals it mostly dominates the budget. Only on large scales does sample variance add to correlations in this case; the non-Gaussian terms are generally of minor importance for all GGL covariance elements.

In Fig. 11 the full covariance is plotted with each element coloured according to the term making the largest contribution by absolute value. It is interesting to see that all components dominate the covariance in some parts, with the exception of the in-survey non-Gaussian term which never contributes more than ~ 10 % to any element (the same conclusion was also reached by Barreira et al. 2018a). While super-sample covariance is critical to include in the cosmic shear covariance, especially for correlations involving low-redshift source samples (top left part of the cosmic shear block) where the survey volume is smaller, it is the three Gaussian components whose accuracy drives the overall fidelity of the model. These terms are primarily assessed through the comparison with our mocks (see below). The importance of precise shear calibration is also evident: smaller uncertainty on the multiplicative bias has the potential to de-correlate most elements of the data vector and thereby improve constraining power.

5.3. Comparison with simulated covariances

The Flask + Salmo mocks are designed for high accuracy in the galaxy sample population properties as well as the two-point
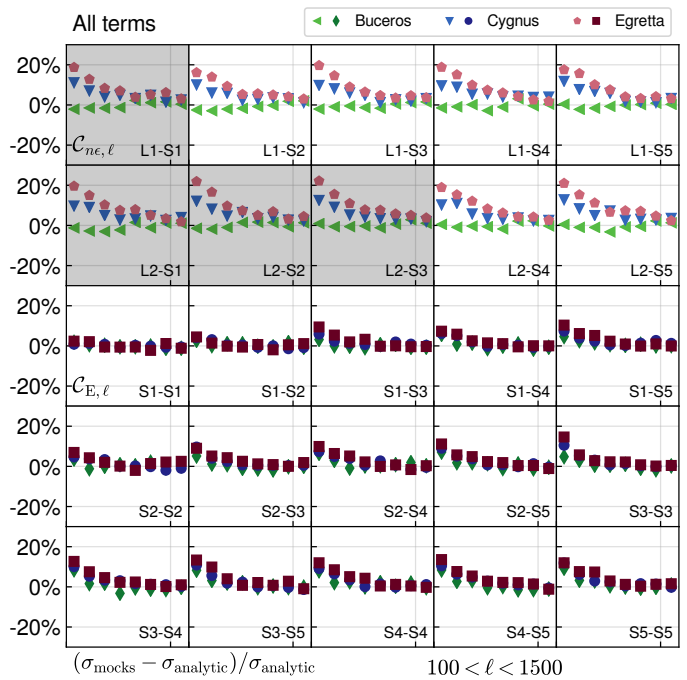
statistics of the weak lensing signal, so that they are well suited to validate the Gaussian terms in the analytic covariance. Due to the lognormal approximation of the matter density distribution in the simulations, the connected four-point function underlying the non-Gaussian covariance contributions is less accurate, which is tolerable as these terms are overall subdominant. Nonetheless, the lognormal assumption is capable of reproducing all salient features in the sample covariance of weak lensing correlation functions (Hilbert et al. 2011). It should be noted that, while Hilbert et al. (2011) assumed lognormality for the weak lensing convergence, we make the more accurate assumption of a lognormally distributed matter density field, which is known to be a very good approximation to the one-point probability density, as well as the third and fourth moments, of the matter density contrast (Kayo et al. 2001).

The first point we address is how the combination of the realistic, patchy survey footprint of KiDS-1000 and the spatial variations in the survey depth affect the covariance. This is shown in Fig. 12 where we plot the relative difference between the Egretta and Buceros setups[16]. Compared to a simple, uniform survey of identical area, the square root of the diagonal of the covariance increases by about 20 %, decreasing smoothly to ~ 10 % excess in the highest source bin combinations for cosmic shear. This trend is driven by shape and shot noise through modifications of the number of galaxy pairs available in the correlation function measurement; super-sample covariance modifies the ratio in the opposite direction but remains subdominant on the diagonal.

---

[16] In all comparisons of this section we only include the covariance contributions of cosmological origin, i.e. there is no multiplicative shear bias calibration uncertainty.
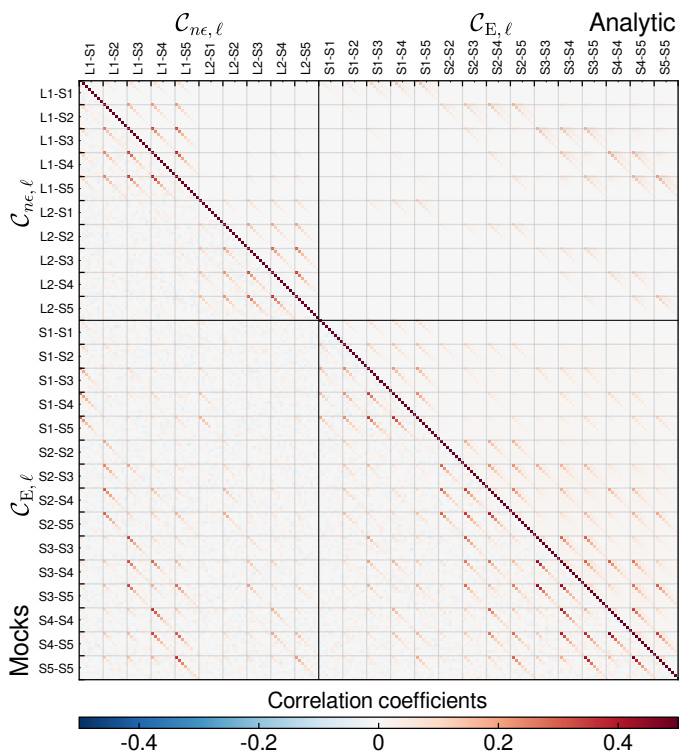
Fig. 14. Comparison of correlation coefficients of the weak lensing band power covariance between mocks (lower left) and analytic covariance (upper right) in the most realistic setup (Egretta). The top left block is for GGL; the bottom right block for cosmic shear. Galaxy sample combinations are indicated on the axes.

The analytic model accurately captures the majority of the survey property effects as it employs the pair counts measured from the real data. The only substantial deviation occurs for GGL on large scales where the analytic approach predicts much smaller changes. We find that this discrepancy is caused by the mixed noise-sample variance term $\mathrm{Cov}_{\mathrm{G,mix}}$ which in the analytic model disregards survey geometry effects altogether.

An alternative view is provided by Fig. 13 in which we directly compare mock and analytic covariances for the three survey setups Buceros, Cygnus, and Egretta. The agreement for cosmic shear is generally very good and independent of the survey setup. At small angular frequency the analytic model underpredicts the standard deviation by typically 10 %, especially at higher redshift, which again originates from the idealised treatment of survey geometry in the Gaussian sample variance terms. We find these limitations to also be behind the interesting trends that emerge in the GGL parts of the covariance: mocks and theory match nearly perfectly in the case of Buceros, but deviate on large scales when switching to the realistic footprint (Cygnus) and a little further when introducing variable depth (Egretta), reaching +20 %.

We compare the off-diagonal terms via the correlation matrix shown in Fig. 14, for the most realistic (Egretta) survey setup. Generally, the correlations between the weak lensing band powers are small, not exceeding 0.5. Mock and analytic model agree to a high degree in the structure of the correlation matrix with the vast majority of elements differing in agreement with the statistical uncertainty due to the finite number of mocks. As the only clear systematic pattern, the sub-diagonals in the cross-covariance between GGL and cosmic shear are more pronounced in the mock covariance. Consulting Fig. 11, these are terms that

are also dominated by the mixed term, thus pointing to the aforementioned limitations once again.

In Appendix D we provide a more detailed discussion of mock and analytical covariances, including the performance of individual terms and results for the corresponding configuration-space signals. In Sect. 7 we investigate the impact on the width of posterior parameter constraints of residual differences in our covariance models, due to different levels of realism in accounting for survey properties as well as due to limiting assumptions in our analytic model.

However, one could also incur a bias in the central value of the posterior if the covariance in the likelihood does not accurately describe the noise properties of the data vector. For a rough estimate of the magnitude of such bias in our analysis, we calculate the systematic difference between data vectors generated with noise from two different covariances labelled A and B. Assuming a Gaussian likelihood, these noisy data vectors of size N_d are given by $\boldsymbol{d}_i = \boldsymbol{\mu} + \mathsf{L}_i \boldsymbol{z}$ with $i \in \{\mathrm{A, B}\}$, where $\boldsymbol{z}$ is a random variate drawn from a multivariate standard normal distribution (with zero mean and unit covariance). We denote the mean of the data vector by $\boldsymbol{\mu}$, and L is a triangular matrix obtained from the Cholesky decomposition of the covariance matrix, $\mathrm{Cov} = \mathsf{L}\mathsf{L}^\tau$.

We now consider the difference between two data vectors generated with the same random variate but different covariances, $\Delta \boldsymbol{d} := \boldsymbol{d}_\mathrm{B} - \boldsymbol{d}_\mathrm{A}$. Since the mean is shared, it cancels in the difference $\Delta \boldsymbol{d}$. The expected variance of this difference is then given by

$$\sigma^2_{\Delta\mathrm{Cov},\alpha} := \left\langle \Delta d^2_\alpha \right\rangle = \sum_{i=1}^{N_\mathrm{d}} \left(L_{\mathrm{B},\alpha i} - L_{\mathrm{A},\alpha i} \right)^2 \ , \tag{42}$$

for an element of the data vector indexed by α, where we exploited the relation $\left\langle z_\alpha z_\beta \right\rangle = \delta_{\alpha\beta}$. For the example of analytic covariances in the Buceros and Egretta setups as models A and B, we obtain for the relative shift in the weak lensing signal, that is $\sigma_{\Delta\mathrm{Cov},\alpha}/\mu_\alpha$, values between 4×10^{-3} and 10^{-4} for GGL and values smaller than 5×10^{-5} for cosmic shear. From this we conclude that biases due to remaining inaccuracies in the covariance model are well below the sensitivity of the current generation of surveys.

6. Parameter inference methodology

We build a dedicated inference pipeline[17] using the CosmoSIS[18] (Zuntz et al. 2015) analysis framework, assembling a mix of its standard components and new bespoke modules. We choose the nested sampling algorithm MultiNest[19] (Feroz & Hobson 2008; Feroz et al. 2009, 2019) to explore the resulting posterior distributions, as we find it to strike a good balance between computational speed and robustness in navigating high-dimensional parameter spaces with several highly non-linear degeneracies.

6.1. Parameters and priors

In Table 3 we list our fiducial set of model parameters and their priors, with a total of 20 parameters varied for joint clustering and weak lensing analyses. The five free parameters describing a spatially flat ΛCDM cosmology are largely adopted from

---

[17] The KiDS Cosmology Analysis Pipeline (KCAP) will be made public on acceptance of the KiDS-1000 analysis papers.
[18] https://bitbucket.org/joezuntz/cosmosis/wiki
[19] https://github.com/farhanferoz/MultiNest

Table 3. Choice of fiducial model parameters and priors.

Parameter	Symbol	Prior
CDM density	ω_c	[0.051, 0.255]
Baryon density	ω_b	[0.019, 0.026]
Density fluctuation amp.	S_8	[0.1, 1.3]
Scalar spectral index	n_s	[0.84, 1.1]
Hubble constant	h	[0.64, 0.82]
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 [5]	$\mathcal{N}(\mathbf{0}; C_{\delta z})$

Notes. The first section corresponds to the primary cosmological parameters. The second section contains astrophysical parameters related to galaxy bias, intrinsic alignments, and baryon feedback on the matter power spectrum. The third section lists observational nuisance parameters. Prior values in square brackets are the limits of top-hat priors, while $\mathcal{N}(\mu; C)$ corresponds to a normal prior with mean μ and (co-)variance C (cf. Fig. 7). The number in square brackets indicates how many instances of the parameter enter the analysis.
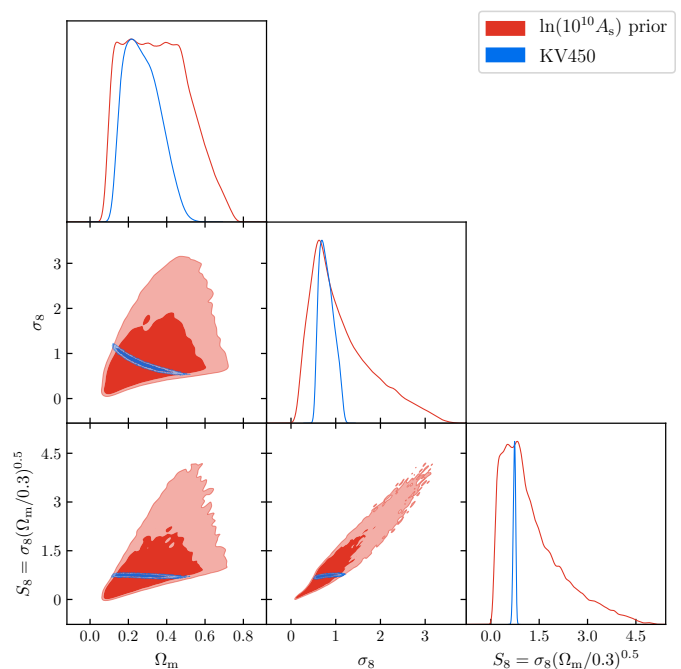

Fig. 15. Prior volume resulting from using wide top-hat priors on the sampling parameters $\{\ln A_s, \omega_c, \omega_b, n_s, h\}$ (in red), shown together with the KV450 posterior (in blue). Dark (light) shades correspond to the two-dimensional 1σ (2σ) credible regions. The prior does not impact on marginal S_8 constraints, but is informative in the $\Omega_m - S_8$ and $\Omega_m - \sigma_8$ planes.

CMB analyses for which they are natural, which means that they are approximately independently constrained with clear links to features in the observables (Kosowsky et al. 2002). This is not necessarily true for low-redshift probes of large-scale structure. In particular, weak lensing is primarily constrained through the amplitudes of its otherwise smooth two-point signals, described by a non-linearly degenerate combination of Ω_m and σ_8. It has therefore become customary to report constraints from cosmic shear in terms of the parameter $S_8 = \sigma_8 (\Omega_m/0.3)^{1/2}$, approximately measuring the lateral width of the posterior across the aforementioned degeneracy. We note however that the exact direction of the $\Omega_m - \sigma_8$ degeneracy is affected by the source redshift range of the data set, as well as the range and weighting of angular scales entering the likelihood.

Standard sampling parameters for the CMB, like the amplitude of the primordial power spectrum of scalar density fluctuations, A_s, and the physical cold dark matter density parameter, ω_c, remain largely unconstrained individually. The choice of priors for these parameters therefore directly impacts on the posteriors of weak lensing cosmological analyses, as illustrated in Figs. 15 and 16. They show posteriors of the key cosmological parameters Ω_m, σ_8, and S_8 for the cosmic shear analyses of KV450 and DES Year 1 (Troxel et al. 2018b), respectively. Alongside, we plot the corresponding prior volumes resulting from the cosmological sampling parameters chosen in each case, mapped by drawing 500 000 samples. The analyses assumed wide top-hat priors for the standard flat ΛCDM cosmological parameters, which we adopt here; in particular, KV450 used $\ln(10^{10} A_s)$ as a sampling parameter with prior range [1.7, 5], while DES Year 1 used A_s in the range $[5 \times 10^{-10}, 5 \times 10^{-9}]$. These assumptions translate into complex prior volumes for the derived low-redshift parameters. Figures 15 and 16 highlight that the extent of the $\Omega_m - \sigma_8$ degeneracy, and thus the marginal posterior width for Ω_m, are almost entirely determined by the prior and not influenced significantly by the different constraining powers of the data sets (see also Joudaki et al. 2020 for a discussion of the impact of the different A_s priors).

In light of the observed tension in S_8 constraints from *Planck* and galaxy weak lensing probes it is particularly important

to avoid implicit informative priors on this parameter. Indeed, Figs. 15 and 16 highlight that both the KV450 and DES Year 1 setups disfavour high values of $S_8 \gtrsim 1$ a priori. However, the S_8 marginal posteriors remain unaffected by the prior as both data sets are highly constraining, and the S_8 prior is flat across the region with substantial posterior mass[20]. To guard against any undue impact by the prior choice and to simplify the parameter volume that needs to be explored by the sampler, we choose S_8 as our sampling parameter instead of A_s or $\ln A_s$ and impose a top-hat prior that comfortably encompasses the full range of the previous, less constraining weak lensing analyses, as well as the *Planck* S_8 posterior. We further assess the impact of the choice of fluctuation amplitude parameter for sampling in Sect. 7.

The other cosmological sampling parameters remain the same as in KV450, with unchanged wide priors for ω_b and h. We tighten the prior range for n_s in order to avoid regions of very small expected posterior mass at large values of n_s, where evaluations of the S17 clustering likelihood become prohibitively slow. To avoid prior volume artefacts in this weakly constrained parameter, we then symmetrise the top-hat prior around the theoretical expectation of values just below unity and set $n_s \in [0.84, 1.1]$.

Since the prior for ω_c directly impacts the inference on Ω_m and σ_8, we seek to derive a more physically motivated prior range than the very wide $\omega_c \in [0.01, 0.99]$ chosen in KV450. We opt to motivate the prior from the independent measurement of luminosity distance via Supernova Type Ia measurements with the Pantheon sample (Scolnic et al. 2018). Their analysis yielded $\Omega_m = 0.298 \pm 0.022$ for a flat ΛCDM model, which we translate into the top-hat prior range $\Omega_m \in [0.188, 0.408]$ that en-

---

[20] By probability mass, we refer to the probability density function (PDF) integrated over certain regions of the PDF's domain.
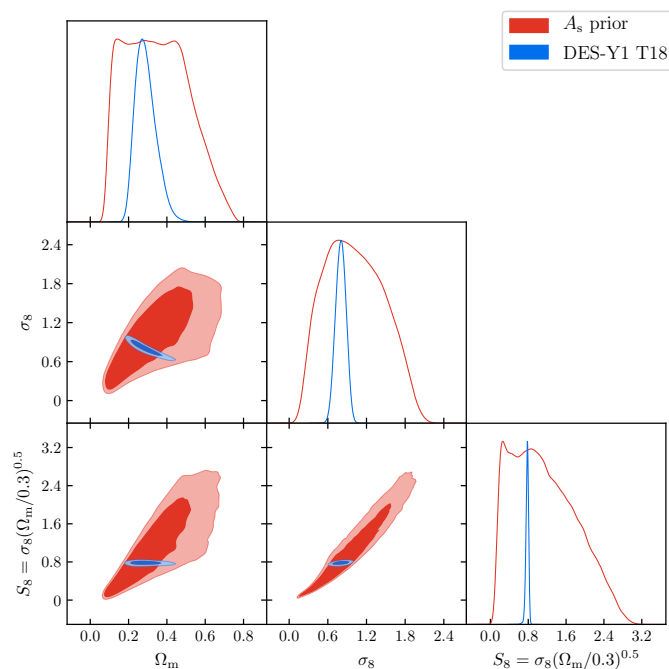
Fig. 16. Same as Fig. 15, but now showing the prior volume using the sampling parameter A_s with a wide top-hat prior, alongside the DES Year 1 cosmic shear posterior. It should be noted that the axis range for the σ_8 and S_8 parameters is smaller than in Fig. 15.

compasses the $\pm 5\,\sigma$ uncertainty of the marginal constraint including systematic errors. The boundaries of the Ω_m prior are then translated into those for ω_c using the extremes of the prior ranges of ω_b and h. This choice also has the benefit of excluding parameter combinations for which the matter power spectrum model underlying all large-scale structure probe predictions is essentially untested – for instance, the Coyote Universe (Heitmann et al. 2014) suite of simulations used to calibrate the Mead et al. (2015) non-linear model is limited to the ranges $\Omega_m h^2 \in [0.12, 0.155]$ and $\sigma_8 \in [0.616, 0.9]$.

The priors for the astrophysical parameters are adopted from KV450 and Tröster et al. (2020b) respectively for IA and galaxy bias. The intrinsic redshift dependence of the IA signal is by default kept fixed at $\eta_{IA} = 0$. We follow S17 in tying the otherwise poorly constrained non-local bias parameter γ_2 to the linear bias,

$$\gamma_2 = -\frac{2}{7}\,(b_1 - 1)\ . \tag{43}$$

This relation is physically motivated and holds to sufficient accuracy in realistic simulations. While a similar relation exists for γ_3^-, it is varied in the inference as a catch-all parameter for other non-linear power spectrum contributions not explicitly modelled in the perturbative approach. Finally, we include one offset parameter δ_z each for S1-5 that shifts the source redshift distributions within the multivariate Gaussian prior discussed in Sect. 3.3.

6.2. Gaussian likelihood assumption

Along with the vast majority of large-scale structure cosmological analyses, we adopt a multivariate Gaussian likelihood. This is expected to be a generally excellent approximation if the summary statistics entering the likelihood have been averaged over
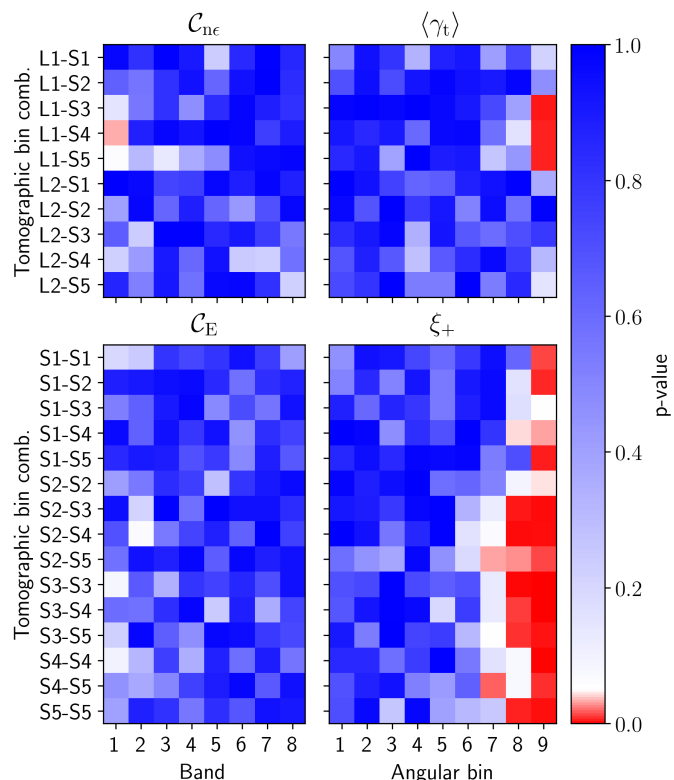

Fig. 17. p-values for a KS-test of Gaussianity of the marginal likelihood distributions. Values above the typical threshold of 0.05 are blue; values below are red. Left panels show band powers for cosmic shear (bottom) and galaxy-galaxy lensing (top) for various tomographic bin combinations and angular frequency bands, with the largest scales on the left. Right panels display corresponding real-space correlation functions, with the largest scales on the right.

many modes in the underlying fields. Exact likelihood expressions for two-point statistics of the non-Gaussian matter distribution and its tracers are unknown, and since only the central value and the covariance are usually readily calculable, a normal distribution is the least informative form of likelihood. Due to the small number of modes entering power spectra at low angular frequencies, significant deviations from Gaussianity occur (e.g. Hamimeche & Lewis 2008), which propagate into configuration-space statistics that are sensitive to these modes, such as the cosmic shear correlation function ξ_+ at large angular separation (Schneider & Hartlap 2009; Sellentin et al. 2018; Lin et al. 2020).

It is therefore an additional benefit of band powers that they have sensitivity over a compact range of angular frequencies, allowing us to reduce the impact of scales below $\ell \lesssim 50$ to a minimum; see Fig. 2. To demonstrate the impact, we create sampling distributions of both band power and configuration-space statistics measured from 1000 mock realisations. We test the one-dimensional likelihoods of the elements of our data vector individually for Gaussianity, applying a Kolmogorov-Smirnov (KS) test to the standardised distributions and setting a threshold p-value of 0.05. As shown in Fig. 17, the band power likelihoods are consistent with being Gaussian, with a slight trend towards smaller p-values for the low-ℓ bands. In contrast, the correlation function likelihoods (in the standard KV450 setup with 9 logarithmically spaced bins in the range from $0.5 - 300$ arcmin) fail in the two to three largest scale bins for ξ_+ and to a lesser degree in γ_t. We note in passing that our test is necessary but

not sufficient because the multivariate distribution of band powers could be non-Gaussian despite Gaussian marginals. In future one could additionally apply tests that are sensitive to higher-order moments of the distribution (see e.g. Sellentin & Heavens 2018).

As we can proceed with a Gaussian likelihood for the band-power statistics, it is important to emphasise that the covariance should not be varied alongside the mean when sampling the parameter space. It is a fundamental property of the normal distribution that its sample mean and covariance are statistically independent. The fact that in large-scale structure analyses both the mean (i.e. the model) and the covariance depend on cosmological and other parameters (primarily through sample variance) is a relic of the true, non-Gaussian sampling distributions that we approximate. Varying parameters in the covariance in the posterior sampling of a Gaussian likelihood would generate spurious, albeit small, extra constraining power because the parameter sensitivity of the covariance is treated erroneously as an independent source of information (see the discussion in Carron 2013).

If the parameters are not varied in the covariance, the question arises which combination of parameters is the correct one to choose, as particularly amplitude-changing parameters like σ_8 or galaxy bias can modify sample variance contributions considerably (e.g. Harnois-Déraps et al. 2019). van Uitert et al. (2018) used an iterative approach to solve the issue, starting with a fiducial set of parameters and then repeating the likelihood analysis with the covariance re-calculated at the best-fit values of the previous iteration. In a forthcoming publication we will demonstrate that this procedure robustly converges to the correct posterior. Since we have best-fit values from the Tröster et al. (2020b) analysis for guidance, we expect that at most one iteration will be required for convergence on the key cosmological parameter constraints. For the mock analysis in this work, we evaluate the covariance at the fiducial parameter values.

If a mock-derived covariance is used in the likelihood, we correct for noise bias in the mean of the inverse covariance (Kaufman 1967; Hartlap et al. 2007). Due to the large number of realisations, the correction is only 4 % and has negligible impact on the posterior.

6.3. Goodness of fit

The goodness of fit is a key statistic to judge how well the combination of the model and the assumed likelihood (including the covariance) describe the data. Traditionally, the weighted least squares, or χ^2, statistic is evaluated at a choice of best fit of the sampling parameters and interpreted under the assumption that it follows a χ^2 distribution with $k_{\rm dof} = N_{\rm d} - N_{\Theta}$ degrees of freedom. Here, $N_{\rm d}$ is again the size of the data vector $\boldsymbol{d}$ entering the likelihood, varying between 120 for cosmic shear only and 310 (adding 168 for clustering and 22 for GGL) for the joint analysis in their default setup, and N_{Θ} is the number of sampling parameters, where $N_{\Theta} = 12$ for cosmic shear and $N_{\Theta} = 20$ for the joint analysis. The goodness of fit is then quantified by the probability to exceed (PTE) the value of χ^2 at the best fit, that is $\Pr(\chi^2 > \chi^2_{\rm best})$.

The assumption of the particular form of χ^2 test statistic is only valid if (1) the underlying data is normally distributed, (2) the sampling parameters enter the model linearly, and (3) the parameter ranges are not significantly restricted by priors. Neither of these assumptions generally hold for cosmological analyses, and ours is no exception. We therefore need to establish a more accurate form of the sampling distribution of the best-fit χ^2. This
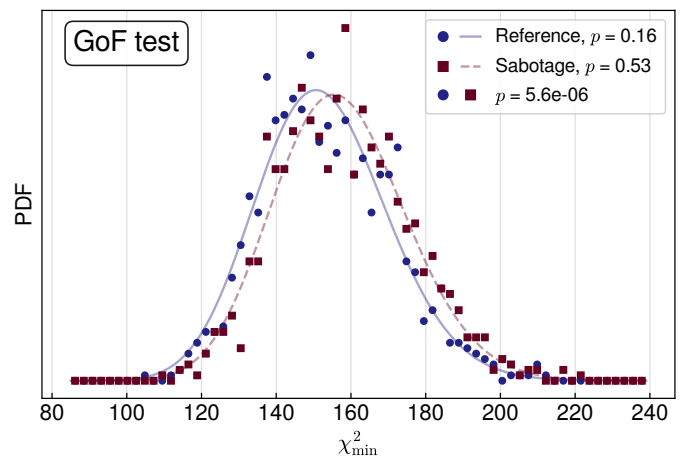
Fig. 18. Goodness of fit sampling distributions. Shown are histograms of minimum χ^2-values obtained from 1000 mock realisations of data vectors analysed with our cosmic shear and GGL analysis pipeline. Blue points correspond to our baseline setup; red points to the result using a sabotaged data vector computed with the S3 and S4 redshift distributions systematically shifted upwards by 5σ of their Gaussian prior width. The curves are best fits of a χ^2-distribution with its mean as free parameter. The solid blue curve is a good fit to the baseline results (p-value of 0.16), while the sabotaged distribution is still well fitted by a χ^2-distribution (red dashed curve; $p = 0.53$) which however is incompatible with the baseline ($p < 10^{-5}$).

distribution can be built empirically from our mocks; if mocks are unavailable, realisations of $\chi^2_{\rm best}$ could analogously be generated directly from the data analysis via the posterior predictive distribution (cf. Gelman et al. 1992, 1996).

For every mock (or posterior predictive) realisation we maximise the posterior, or alternatively minimise the log-likelihood, to obtain an estimate of $\chi^2_{\rm best}$. An example of a resulting probability density function (PDF) created from 1000 mocks is shown in Fig. 18 for the case of joint cosmic shear and GGL. We could now directly measure the compatibility of the $\chi^2_{\rm best}$ determined from the real-data analysis with being a sample drawn from this PDF via $\Pr(\chi^2_{\rm best,mock} > \chi^2_{\rm best,obs})$. However, for the joint clustering and weak lensing analysis in particular, we struggled to identify readily available optimisation algorithms[21] that ran fast enough to be applicable to a large number of mock realisations. Fortunately, in the cases where the $\chi^2_{\rm best}$ PDF is obtainable, it is still well described by a χ^2-distribution; see Fig. 18, and it is fair to assume that this remains the case for the most comprehensive analysis setup. The shape of the distribution is retained even for a substantially under-fitting model as shown by the red points and curve in Fig. 18.

Therefore, the problem reduces to finding an estimate of the one free parameter of the $\chi^2_{\rm best}$ sampling distribution, which we denote by $k_{\rm dof,eff} := N_{\rm d} - N_{\Theta,\rm eff}$. This task can be intuitively understood as obtaining an effective number of constrained model parameters, $N_{\Theta,\rm eff}$, which will be smaller than the raw number of parameters due to parameter degeneracies and informative priors that severely restrict the freedom of a parameter to optimise the model. We investigate a number of estimators for $N_{\Theta,\rm eff}$ that

---

[21] We did not explore algorithms that require function derivatives explicitly, such as conjugate gradient approaches, but recommend this for future work.

Table 4. Estimates of the effective number of model parameters.

No. of pars. estimate	Cosmic shear			All WL		
	μ	σ	α_p	μ	σ	α_p
N_Θ	12	-	22.6	18	-	32.8
$N_{\Theta,\mathrm{var}}$	3.0	0.5	−4.0	7.4	1.3	−4.3
$N_{\Theta,\mathrm{like}}$	5.4	1.3	3.2	10.7	1.8	7.2
$N_{\Theta,\mathrm{post}}$	3.5	0.6	−2.3	7.6	1.1	−3.9
none	0	-	−12.6	0	-	−29.4
$N_{\Theta,\mathrm{eff}}$	4.5	0.8	0.4	9.1	1.3	1.6
from $\langle \chi^2 \rangle$	4.5	0.6	0.4	8.7	0.4	0.2

Notes. Rows correspond to the total number of parameters varied in the analysis, N_Θ, the three estimators for the effective number of parameters given in Eqs. (44) to (46), and the assumption of no parameters varied. We also show our adopted estimator of $N_{\Theta,\mathrm{eff}} = (N_{\Theta,\mathrm{like}} + N_{\Theta,\mathrm{post}})/2$ and the effective number of parameters as inferred from the mean of the sampling distribution of minimum χ^2-values (as in Fig. 18). Columns denote the central value μ, the standard deviation σ, and the quality of match to the χ^2_{min} sampling distribution, α_p (cf. Eq. 47).

have seen applications in the astrophysical literature:

$$N_{\Theta,\mathrm{var}} = 2\left[\left\langle\left(\chi^2\right)^2\right\rangle_{\mathrm{Pr}(\Theta|d)} - \left\langle\chi^2\right\rangle^2_{\mathrm{Pr}(\Theta|d)}\right] ; \qquad (44)$$

$$N_{\Theta,\mathrm{like}} = \left\langle\chi^2\right\rangle_{\mathrm{Pr}(\Theta|d)} - \chi^2_{\mathrm{min}} ; \qquad (45)$$

$$N_{\Theta,\mathrm{post}} = \left\langle\chi^2\right\rangle_{\mathrm{Pr}(\Theta|d)} - \chi^2\left(\mathrm{argmax}\left[\mathrm{Pr}(\Theta|d)\right]\right) , \qquad (46)$$

where $\mathrm{Pr}(\Theta|d)$ denotes the posterior, and angular brackets indicate an average over the posterior distribution (or samples thereof). Equation (44) was recently proposed by Handley & Lemos (2019a). The other two expressions are variants of the basic estimator introduced by Spiegelhalter et al. (2002), which employs a point estimate of χ^2 at a point where 'information is maximised'. We experiment with the maximum likelihood (Eq. 45) and the mode of the posterior (Eq. 46; see Raveri & Hu 2019 for a recent application), the former attaining the global minimum χ^2 and the latter more attentive to the informative priors in our analysis. The posterior mean is a potential alternative (Kunz et al. 2006), but as it would always yield lower values of $N_{\Theta,\mathrm{eff}}$ than Eq. (46), it does not perform favourably in our setup.

Figure 19 shows the performance of the estimators in relation to $N_d - k_{\mathrm{dof,eff}}$ as inferred from the χ^2_{min} sampling distribution (cf. Fig. 18). Neither estimator is a good match per se. Since $N_{\Theta,\mathrm{var}}$ avoids a point estimate, it is the least noisy, but also furthest from the target. We saw hints that the accuracy of this estimator is significantly affected by the presence of informative priors. The other two estimators bracket the target in both scenarios that we can assess, and since neither is a priori more preferable, we resort to define our final estimate as their average, $N_{\Theta,\mathrm{eff}} := (N_{\Theta,\mathrm{like}} + N_{\Theta,\mathrm{post}})/2$.

Table 4 quantifies the performance of the various parameter-number estimates, listing their central values, standard deviations, and the quality measure

$$\alpha_p := \frac{1 - 2\mu_p}{\bar{\sigma}_p} , \qquad (47)$$

with μ_p and $\bar{\sigma}_p$ the mean of the p-values and its standard error calculated for a fit of a χ^2-distribution with $k_{\mathrm{dof,eff}}$ degrees of freedom to the χ^2_{min} sampling distribution. Since the p-value is expected to follow a uniform distribution in $[0; 1]$, the expectation value of μ_p is 0.5 and its standard error is $1/\sqrt{12}$, so that the

error on the mean is $\bar{\sigma}_p \approx 0.03$ for the 100 realisations that we use. Smaller absolute values of α_p are preferable, with $|\alpha_p| < 1$ indicating that the bias of the estimator is subdominant to its statistical error.

Our choice of estimator performs well in this regard with $\alpha_p \lesssim 1$, while the other effective parameter number measures with absolute values of α_p between 2 and ~ 4, and the traditional N_Θ with $\alpha_p \approx 23$ and 33 for cosmic shear and the combined weak lensing signals, respectively, are strongly disfavoured. The latter performs even slightly worse than assuming no free parameters at all. However, it should be noted that the $N_{\Theta,\mathrm{eff}}$ is quite noisy if determined from a single maximum likelihood estimate, so that combining at least a few realisations (from mocks or the posterior predictive distribution) is recommended.

It is interesting to note that the effective number of constrained parameters, defined as the equivalent dimension of a linear, unconstrained parameter space, is 4.5 for cosmic shear and approximately 9 for cosmic shear and GGL combined. The classic approach to goodness of fit, employing N_Θ to calculate the reduced χ^2, thus yields a conservative test statistic if model under-fitting is the primary concern. This means that previous cosmological analyses with similarly structured model parameter spaces tended to report reduced χ^2 values whose expectation was slightly larger than unity. The change due to using $N_{\Theta,\mathrm{eff}}$ is mild for our fiducial KiDS-1000 setup (e.g. 7 % in the reduced χ^2 for cosmic shear only), but could become important for a smaller data vector, for instance if data compression is applied.

6.4. Reporting parameter constraints

In cosmology it is widely accepted to adopt the Bayesian paradigm for inference and provide the posterior distribution, usually in the form of samples, as the final deliverable of an experiment. Nonetheless, point estimates of some notion of best-fit parameter value are indispensable to report headline results or compare with other experiments at a high level. Moreover, researchers are often interested in the recovery of a 'true' underlying value of model parameters. This is perhaps most obvious in the validation of an analysis pipeline – we seek to demonstrate that we faithfully recover the input parameters of the mock data.

Usually, point estimates are extracted from the one-dimensional marginal distributions of the parameters (as this low-dimensional distribution is well sampled), typically the mode, median, or mean combined with a credible interval that encompasses a defined fraction of the highest marginal posterior density. While these point estimates are unambiguous in their Bayesian interpretation, they do not necessarily peak at, or even cover within a given credible interval, the true parameter value[22]. In our case this is a consequence of a high-dimensional parameter space with multiple, non-linear near-degeneracies, in addition to a wide prior with complex shape. Figure 20 displays a pertinent example: the strong, banana-shaped degeneracy between Ω_m and σ_8 present in the posterior of cosmic shear analyses.

We therefore seek to employ a complementary point estimate that more accurately reports the global best fit to the data at hand. The estimate itself is simply the set of parameter values at the maximum (multivariate) posterior, denoted by MAP (maximum a posteriori):

$$\Theta_{\mathrm{MAP}} = \mathrm{argmax}\,\mathrm{Pr}(\Theta|d) , \qquad (48)$$

---

[22] Pathological examples with arbitrarily low marginal posterior density at the position of the maximum of the joint posterior are straightforward to construct.
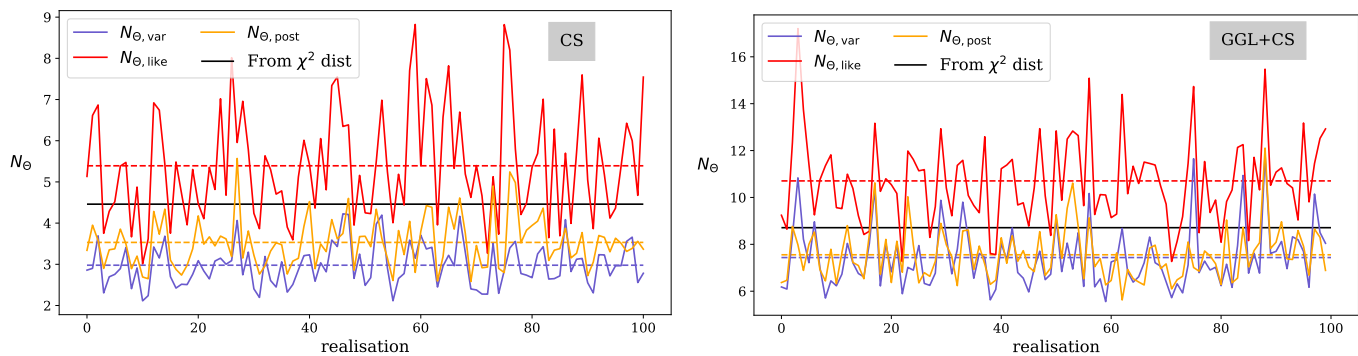
Fig. 19. Estimates of the effective number of parameters for 100 realisations of a cosmic shear *(left)* and a joint cosmic shear+GGL *(right)* data vector. The estimators defined in Eqs. (44) to (46) are shown in blue, red, and orange, respectively. The dashed lines mark the average over the realisations. The black line marks the effective number of parameters inferred from the fit of a χ^2-distribution to a histogram of $\chi^2_{\min}$ (cf. Eq. 47).

where argmax returns the argument of a function at which it attains its maximum. The posterior maximisation advocated in Sect. 6.3 for determining the goodness of fit also provides a precise MAP estimate. In practice, we find good recovery of the global maximum by taking the largest posterior returned by a suite of tens of optimisation runs, each started at the position of a different posterior sample with high probability mass.

The decision-theoretical optimal uncertainty estimate for the MAP is the credible region $\mathcal{H}_\alpha$ defined by the highest posterior density (Robert 2001; see also Price et al. 2020 for another recent application in cosmology)

$$\mathcal{H}_\alpha = \{\boldsymbol{\Theta} : \Pr(\boldsymbol{\Theta}|\boldsymbol{d}) > c_\alpha\} , \qquad (49)$$

for some threshold c_α, depending on α, the posterior mass enclosed within $\mathcal{H}_\alpha$:

$$\int_{\boldsymbol{\Theta} \in \mathcal{H}_\alpha} \mathrm{d}^{N_\Theta}\boldsymbol{\Theta} \, \Pr(\boldsymbol{\Theta}|\boldsymbol{d}) = \alpha . \qquad (50)$$

For our purposes, the definition Eq. (49) is not practical, however. We wish to use a credible interval for the MAP estimate that can be easily compared to established uncertainty estimates, such as the standard deviation of posterior samples or the highest posterior density in the marginal distribution of a given parameter (denoted by M-HPD hereafter). While straightforward in one dimension, this is not the case for the high-dimensional parameter spaces we are considering here. For example, while the notion of 1σ credible intervals readily translates to $\alpha \approx 0.68$ in one dimension, this is not the case for higher dimensions, where α rapidly decreases with increasing number of dimensions.

We propose a hybrid credible interval estimator that is based on the joint, multi-dimensional highest posterior density region, but projected onto the marginal posterior of the parameter under consideration (PJ-HPD henceforth, for projected joint highest posterior density). For each parameter i, define the credible interval as the highest posterior density region that encompasses a fraction α of its marginal posterior mass:

$$\mathcal{H}_i^{\mathrm{marg.}} = \{\boldsymbol{\Theta} : \Pr(\boldsymbol{\Theta}|\boldsymbol{d}) > c_i\} , \qquad (51)$$

such that

$$\int_{\Theta_i \in \mathcal{H}_i^{\mathrm{marg.}}} \mathrm{d}\Theta_i \, \Pr(\Theta_i|\boldsymbol{d}) = \alpha , \qquad (52)$$

where $\Pr(\Theta_i|\boldsymbol{d})$ is the posterior of Θ_i marginalised over all other parameters of the model. In general, this does not correspond
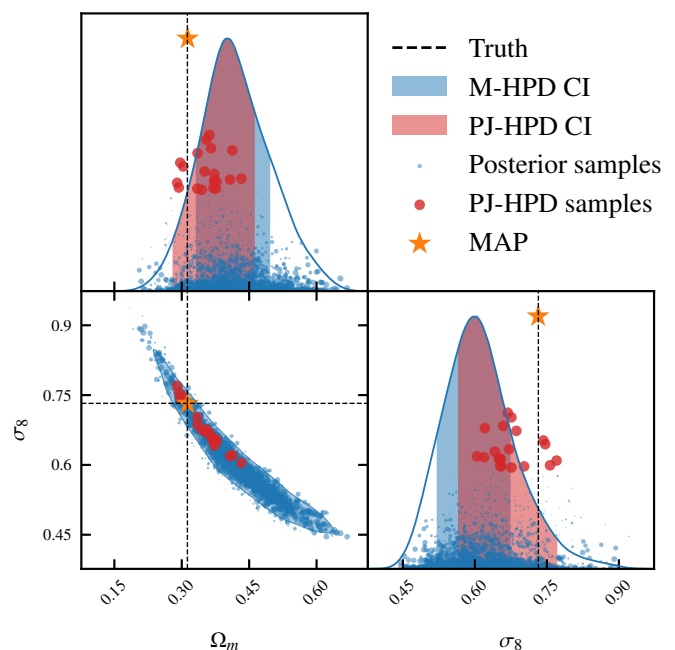
Fig. 20. Illustration of 1σ credible intervals (CI) used to report parameter constraints, for a subset of two parameters for an exemplary mock joint analysis of GGL and cosmic shear with a noiseless data vector. The standard marginal highest posterior density (M-HPD) CIs are shown in blue. The marginal posterior mode is shifted with respect to the true input parameter values (black dashed lines), whereas the maximum a posteriori (MAP) estimate (orange star) tracks the truth well. The CI we associate with the MAP is constructed via the PJ-HPD method (red), using the samples with the highest posterior density (red points). Other posterior samples are shown as blue points, with point size proportional to their posterior mass. In the top and right panels the height of the points and of the MAP indicate their posterior density with arbitrary scaling.

to the region of the highest marginal posterior. The PJ-HPD estimate is guaranteed to include the MAP (since by definition MAP lies within the highest posterior density region) and readily allows for the comparison of the credible intervals between analyses with a different dimension of parameter space. It is intuitive in that the PJ-HPD reduces to the standard M-HPD credible interval in one dimension. Moreover, in the case of a multivariate Gaussian posterior, the PJ-HPD credible interval coincides with the marginal 1σ interval.

To compute our PJ-HPD credible intervals, we take the posterior samples $\{\hat{\mathbf{\Theta}}_n\}$ for $n = 1, 2, \ldots$ from nested sampling chains or MCMC and sort them in decreasing order of their associated posterior density, that is $\Pr(\hat{\mathbf{\Theta}}_{n+1}|\boldsymbol{d}) < \Pr(\hat{\mathbf{\Theta}}_n|\boldsymbol{d})$. Stepping through this list, and for each parameter Θ_i under consideration, we record the interval of parameter values $[\hat{\Theta}_{i,a}; \hat{\Theta}_{i,b}]$ that encompasses all parameter values in the posterior samples up to that point in the list, so $\hat{\Theta}_{i,a} \le \hat{\Theta}_{i,n} \le \hat{\Theta}_{i,b}$ for all n. Then we measure the posterior mass within this interval in the *marginal* distribution of the parameter, $\int_{\hat{\Theta}_{i,a}}^{\hat{\Theta}_{i,b}} \mathrm{d}\Theta_i \Pr(\Theta_i|\boldsymbol{d})$. The iteration stops when the posterior mass in the marginal distribution reaches the desired level α (i.e. usually 68 % or 95 %) upon which the corresponding parameter interval is reported as the PJ-HPD credible interval[23]. In practice, due to the sparse sampling of the highest-density posterior regions, we linearly interpolate the interval boundaries between the samples just before and after the stopping criterion is reached.

This approach is illustrated in Fig. 20 where we show the marginal posterior distributions for the parameters Ω_m and σ_8 out of a total of 18 model parameters sampled in a mock GGL and weak lensing analysis using a noiseless data vector. The one-dimensional marginal posterior modes do not coincide with the input cosmology, whereas the MAP estimate recovers it well (any residual deviation is due to the limited numerical accuracy of the optimisation). While the one-dimensional marginal posteriors have only moderate skewness, the input parameter values and the MAP estimate lie outside the 68 % M-HPD credible interval for both Ω_m and σ_8. In contrast, the PJ-HPD credible interval represents a projection of the highest-density region of the full multivariate posterior that encompasses the MAP.

As the PJ-HPD interval extends to regions of the marginal posterior with lower posterior mass, the resulting constraints are generally slightly less tight than the standard M-HPD credible interval, especially if the posterior deviates strongly from a normal distribution. Moreover, as the PJ-HPD interval is determined from typically tens rather than thousands of posterior samples, its boundaries also scatter more. We find a scatter of order 10 % of the 1σ limits for a multivariate Gaussian that has the same dimension and is sampled with the same number of points as the posteriors in our analysis.

We adopt the MAP+PJ-HPD approach for reporting constraints in the KiDS-1000 analysis alongside the standard marginal posterior mode and associated M-HPD, and employ it as the main criterion for validating our inference pipeline; see Sect. 7. It should be emphasised that any differences between these two credible intervals are *not* an indication of bias in the parameter they report; rather, this occurs if the posterior that they are applied to deviates from multivariate normality.

Finally, we note that posterior sampling techniques have an inherent limit on the accuracy of parameter constraints, dependent on the number of samples drawn. For settings that our computing resources can deliver in reasonable time (always yielding well in excess of 10^4 posterior samples), we find this limit to be $\sim 0.1\sigma$ in the key parameter S_8, corresponding to the scatter of best-fit values between repeat runs of otherwise identical setups. This floor sets a stopping criterion for validation, as well as for

---

[23] We note that the PJ-HPD approach will not necessarily return a simply connected credible interval in the case of a multimodal posterior, nor do we expect it to yield meaningful credible intervals in pathological cases. However, such outcomes should serve as a stark warning not to summarise the inference process via a point estimate in the first place.
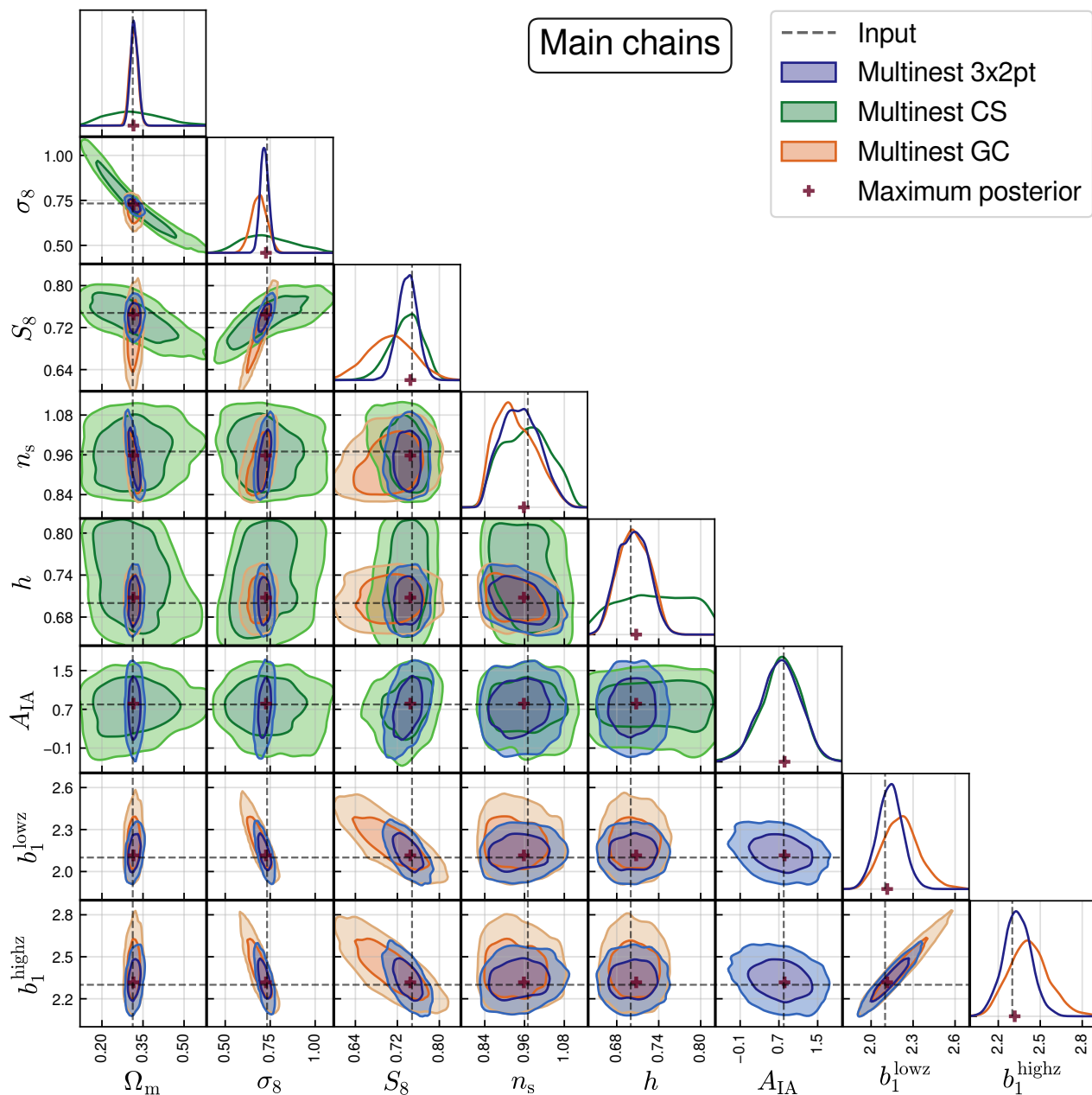
Table 5. Validation tests and resulting S_8 marginal credible intervals.

ID	Setup	M-HPD	PJ-HPD
1	3×2pt, reference	0.037	0.038
2	3×2pt, no GGL overlap	0.037	0.038
3	3×2pt, fiducial	0.037	0.037
4	CS+GGL, reference	0.056	0.062
5	CS+GC, reference	0.036	0.039
6	GGL+GC, reference	0.079	0.108
7	CS, reference	0.061	0.061
8	GC, reference	0.090	0.111
9	CS+GGL, analytic, Buceros	0.054	0.072
10	CS+GGL, analytic, Egretta	0.056	0.062
11	CS+GGL, mocks, Buceros	0.054	0.057
12	CS+GGL, mocks, Egretta	0.057	0.066
13	CS+GGL, S_8 prior, corr δz	0.056	0.062
14	CS+GGL, S_8 prior, uncorr δz	0.057	0.074
15	CS+GGL, $\ln A_s$ prior, corr δz	0.047	0.054
16	CS+GGL, $\ln A_s$ prior, uncorr δz	0.045	0.047

Notes. Setups correspond to those shown in Fig. 22. The reference includes a more progressive selection of scales in the GGL data vector (Bands 1-7) than our fiducial analysis. GC corresponds to galaxy clustering, CS to cosmic shear, and '3×2pt' to the combination of all probes. Analytic/mocks and Buceros/Egretta indicates analytic and simulated covariances with simple or more realistic survey and sample properties. The choice of an (un)correlated prior on the source redshift distribution shifts is labelled as (un)corr δz. GGL signals with strong overlap of the lens and source redshift distributions have been removed in the 'no GGL overlap' case. We note that setup 4 is identical to 13. The two rightmost columns give the width of the 1σ credible intervals (CI) for the standard M-HPD credible interval and the PJ-HPD method, i.e. twice the standard deviation if the posterior was Gaussian.

the mitigation of systematic effects in our data (as reported in Giblin et al. 2020).

7. Validation of the likelihood analysis pipeline

We proceed to validate our inference pipeline by running it on simulated, noiseless data vectors generated from the same modelling pipeline as used in the inference. The absence of noise guarantees that the likelihood peaks exactly at the parameter combination chosen to generate the data vector and is therefore used to demonstrate recovery of the true underlying model parameters. Since we centre our fiducial set of model parameters at the peak of any of the informative Gaussian priors, we also expect the posterior to peak at the input parameter values. As discussed in Sect. 6.4, this will in general not hold for marginal posterior distributions, so we consider the multivariate maximum a posteriori (MAP) estimate.

Runs with noisy data vectors, which have a zero-mean multivariate Gaussian noise realisation drawn with the full covariance added, resemble more closely the processing of the real data. We repeat our validation tests with ensembles of noisy data but find our conclusions identical to the noiseless case. We also create data vectors that systematically deviate from the ideal model in order to assess biases in parameter constraints caused by effects not accounted for in the model. The results of these runs are discussed in the sections where the effect under consideration is covered.

In Fig. 21 we show one- and two-dimensional marginal posteriors of a subset of the 20 sampling parameters that are significantly constrained, alongside projections of the MAP. We also

Fig. 21. Marginal posterior distributions for a selection of parameters that are significantly constrained by the data, including the derived parameters Ω_m and σ_8. Shown are one-dimensional marginals along the diagonal and two-dimensional marginal 1σ and 2σ credible regions elsewhere. Results are for the reference setup using a noiseless mock data vector of galaxy clustering (GC; red), cosmic shear (CS; green), and the combination of clustering and the weak lensing probes ('3×2pt'; blue). Input parameter values are indicated by the black dashed lines, which are excellently recovered by the MAP estimate (red crosses).

include Ω_m and σ_8 as derived parameters. The MAP recovers the input values of all parameters shown in the figure to better than 1.7 %. The combination of clustering and weak lensing signals partially lifts the typical degeneracy between Ω_m and σ_8 for weak lensing-only constraints, through the small clustering errors on Ω_m. This benefits the marginal posterior of the key parameter S_8 as well since in our analysis setup it is still significantly correlated with Ω_m. The joint analysis also leads to visibly tighter and more symmetric marginal posteriors for the linear bias parameters, whereas improvements on the constraints on intrinsic alignments, baryon feedback, and the calibration of the redshift distributions remain marginal.

Figure 22 and Table 5 provide a more quantitative overview of the S_8 constraints for the various probe combinations and analysis setups that we consider (identified by their ID). All credible intervals are extracted from a histogram of the posterior samples smoothed with Gaussian kernel density estimation, with the variance determined by Silverman's rule (Silverman 1986). Both M-HPD and PJ-HPD credible intervals are provided; we use the former to compare S_8 constraints with previous measurements (see below) and recommend the latter to assess the recovery of the input S_8.

The 1σ M-HPD credible interval for cosmic shear alone is 20 % smaller than the KV450 constraint, while the fiducial joint analysis halves the statistical uncertainty on S_8 with respect to
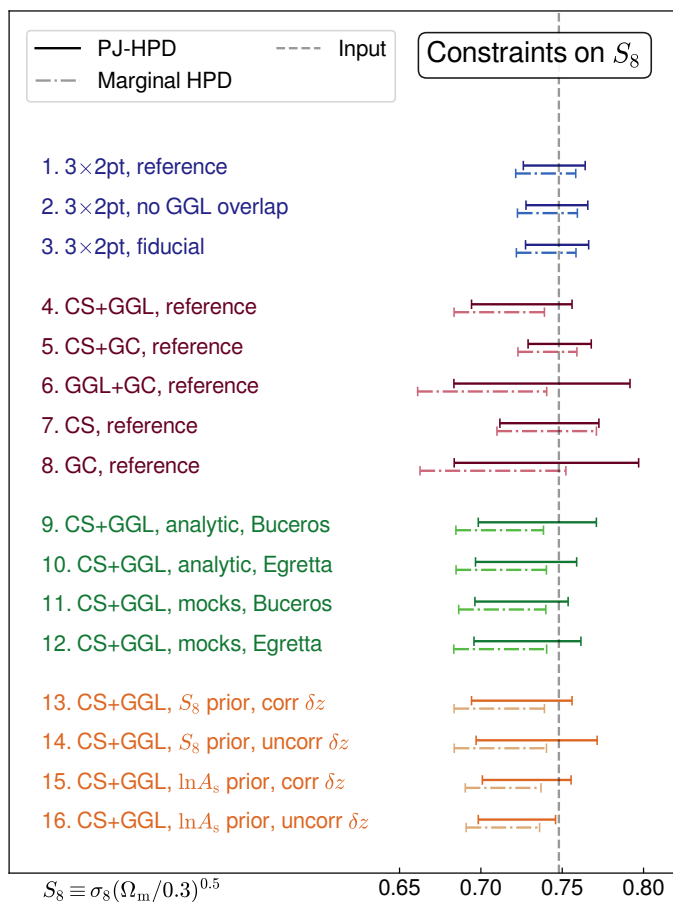
Fig. 22. S_8 constraints (1σ) for different compositions of the GGL data vector in the joint analysis (blue), probe combinations (red), different covariance models (green), and different prior assumptions (orange). Standard marginal (M-)HPD credible intervals are shown as dot-dashed lines; PJ-HPD intervals as solid lines. The grey dashed line indicates the input value of S_8. The corresponding interval widths, as well as explanations of the case labels, are given in Table 5. The M-HPD intervals are generally shifted to S_8 values lower than the input; see text and Sect. 6.4 for a discussion.

KV450 (which corresponds to 29 % smaller errors than the corresponding BOSS+KV450 constraints by Tröster et al. 2020b). The marginal 1σ credible intervals shrink by 10 % (18 %) for Ω_m over BOSS+KV450 (BOSS alone), and by 31 % (40 %) for A_{IA} over BOSS+KV450 (KV450 alone). The precision on linear and non-linear galaxy bias parameters increases by roughly 30 % on most parameters with respect to BOSS only, but most of the added constraining power was already achieved in the Tröster et al. (2020b) analysis, and non-linear bias constraints are still too weak for informative conclusions on galaxy physics.

The constraints on the key parameters S_8 and Ω_m are highlighted in Fig. 23 and put into context with earlier analyses based on 450 deg^2 of KiDS data (Wright et al. 2020b; Tröster et al. 2020b). Both the cosmic shear-only and joint analysis cases are shown, for the reference setup with IDs 7 and 1, respectively. It is evident that in the latter case the tension with *Planck* results will increase if the best-fit remains close to the Tröster et al. (2020b) values, with the S_8 marginal constraints now on par with those originating from the CMB.

Naively, one might expect the KiDS-only parameter errors to decrease by a factor $\sqrt{2}$ because the KiDS-1000 survey area has doubled with respect to earlier KiDS analyses. However,

this does not take into account the highly non-linear parameter dependencies, updated priors, or covariance contributions that do not scale inversely with area, in particular the multiplicative shear bias uncertainty. Most importantly in the cosmic shear case though, for our choices of two-point statistic and scales included in the analysis S_8 is not an optimal summary parameter of the constraining power of the data, as significant correlations with Ω_m remain (see Fig. 23). While a power-law index of 0.5 in the S_8 definition was close to optimal for the $\xi_\pm$ data analysed in KV450, we find that larger values of ~ 0.6 for the index capture the degeneracy between Ω_m and σ_8 better for band powers (this is explored further in Asgari et al. 2020b). The gain in constraining power for an adjusted definition of S_8 is larger and consistent with the naive $\sqrt{2}$ expectation, indicating that the systematic error treatment and the analysis choices have actually evolved in tandem with the statistical power of KiDS-1000.

It is evident that GGL does not add significant constraints to either clustering or cosmic shear, and its addition to their combination has no effect at all on S_8 (cf. Fig. 22 and Table 5, IDs 1 and 4–8). This somewhat sobering finding is caused by the dilution of constraining power due to almost doubling the number of parameters when adding GGL to cosmic shear (by introducing galaxy bias), and due to the more than an order of magnitude larger sky area available for clustering measurements. Accordingly, no palpable difference is seen when cutting the GGL data vector to our conservative, fiducial setup that removes highly non-linear scales and when removing bin combinations with a large fraction of source galaxies in front of lenses (IDs 2,3; see Sects. 2.2 and 3.3). We use the more progressive GGL setting that includes scales with $k \lesssim 1\,h\,\mathrm{Mpc}^{-1}$ (Bands 1–7) for other comparisons to increase chances to detect any issues with the GGL model during validation (hence referred to as the reference). We also refrain from showing GGL-only results as we find that MULTINEST struggles to explore the resulting highly degenerate 18-dimensional posterior, producing unreliable results.

Since we exclude bin combinations with strong overlap of sources and lenses, there is also no self-calibration of intrinsic alignments so that marginal A_{IA} constraints are the same for the joint analysis and cosmic shear alone; see Fig. 21. On the real data the addition of GGL to cosmic shear and clustering does lead to a tightening of the marginal A_{IA} posterior by 17 %, but the marginal posterior maximum also shifts by 12 % (Heymans et al. 2020). Both real and mock analysis agree that GGL does not significant constraining power to the galaxy bias parameters either.

Choosing either an analytic or mock-derived covariance (based on 5000 realisations), with a simple rectangular footprint and uniformly distributed galaxies (Buceros; see Sect. 4.2 for details about the mock setup) or else the realistic survey mask and depth variations (Egretta), has no significant impact on any of the marginal parameter constraints (cf. IDs 9–12)[24]; see also Krause et al. (2017) for analogous conclusions drawn for the DES Year-1 analysis. This highlights that, despite the significant patterns of deviations between the covariance approaches and modelling choices analysed in Sect. 5.3, the weak lensing error modelling is robust at the level of constraining power by KiDS-1000 (or indeed the full KiDS survey, which will add another 35 % of area). It will be interesting to explore in future work how the limitations identified in the covariance model af-

---

[24] For these analyses we kept the δz parameter fixed at 0 to speed up the runs and did not include the uncertainty due to multiplicative shear bias calibration to highlight the differences in the cosmological covariance contributions.
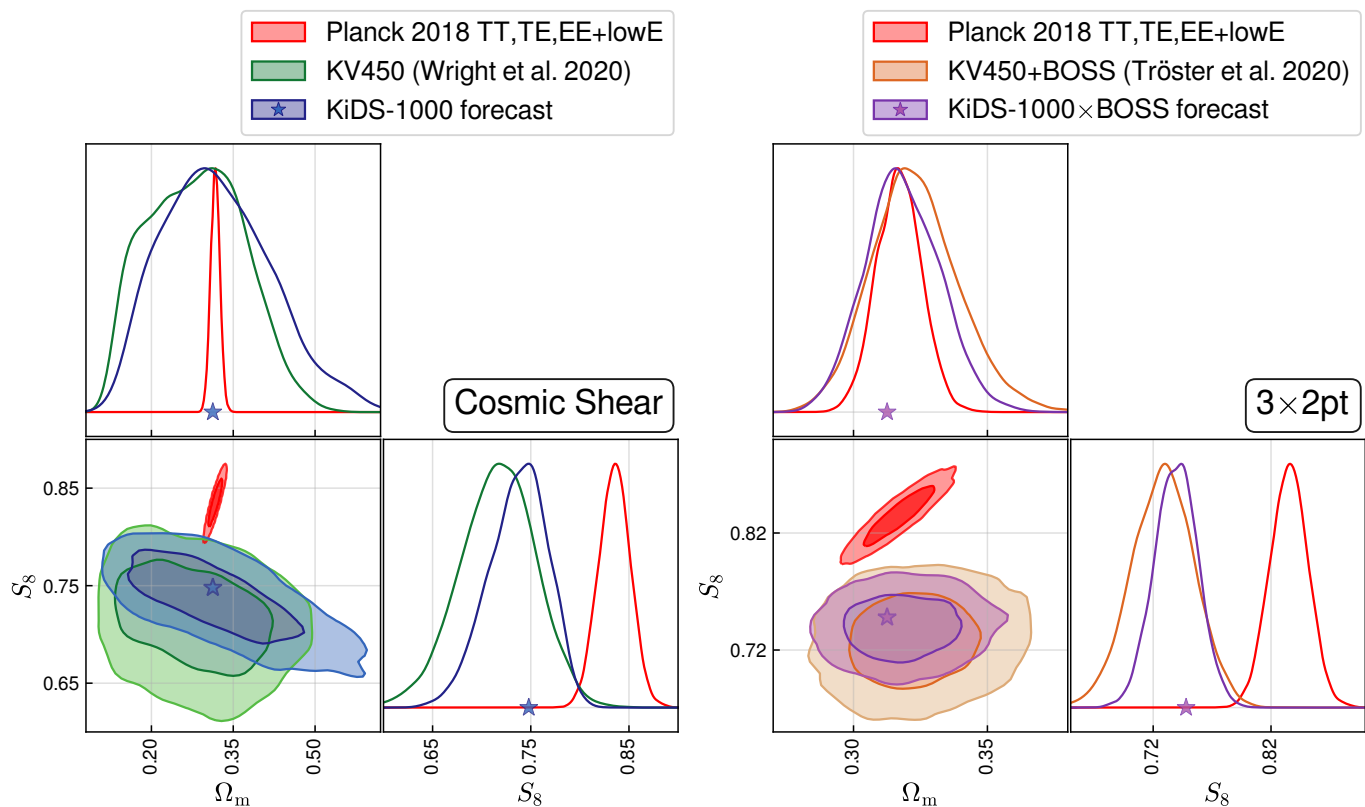
Fig. 23. Predicted KiDS-1000 constraints (reference setup, IDs 1 and 7) on the key cosmological parameters Ω_m and S_8 in relation to previous KiDS results (Wright et al. 2020b; Tröster et al. 2020b) and the Planck Collaboration et al. (2020a) primary CMB constraints, as indicated in the legends. The *left* panel shows results for cosmic shear only; the *right* panel those for the joint analysis of BOSS clustering, galaxy-galaxy lensing between BOSS/2dFLenS and KiDS-1000, and cosmic shear. The star in each panel marks the fiducial cosmology assumed for the forecast. Axis scales differ between the two panels.

fect next-generation surveys that cover substantial fractions of the sky, balancing overall tighter accuracy requirements due to the increased constraining power with a reduced sensitivity to survey boundaries as coverage becomes more contiguous.

The S_8 constraints are also stable under changes from an uncorrelated prior on the source redshift distribution shifts δ_z, to the correlated prior in Fig. 7, as well as under the choice of parametrisation of the normalisation of the matter power spectrum, where we compare $\ln A_s$ with S_8 (cf. IDs 13–16). Our choice of using S_8 directly as the sampling parameter with an associated wide top-hat prior leads to more conservative marginal S_8 constraints, with the 1σ interval extended by 19 % mostly towards lower values. We also consider an additional parameter η_{IA} that allows for extra freedom in the redshift scaling of the IA signals (cf. Eq. 16), imposing a wide top-hat prior in the range ± 5. For a cosmic shear-only analysis this has negligible impact on the cosmological parameter constraints, widening the marginal S_8 posterior by 3 %.

Figure 22 illustrates that the standard M-HPD credible intervals are all skewed towards lower values of S_8 than the input. The displacement is stronger for probe combinations that have weaker constraints on S_8 and leave more pronounced degeneracies in their posterior, which is the case for GGL in combination with either clustering or cosmic shear (IDs 4 and 6). As a consequence, the input S_8 values lies outside the 1σ interval for the cosmic shear+GGL case while the shift is mild in the cosmic shear only (ID 7) and joint clustering and weak lensing cases (e.g. ID 1). The PJ-HPD credible intervals encompass the input value by design (however, it can be right at the interval boundary, or just beyond due to numerical inaccuracy; cf. ID 16) and are generally less displaced but also wider than the M-HPD intervals. This increase in width is minor in the well-constrained cases (negligible for cosmic shear only and 3 % for the joint analysis), but significant in the probe combinations with strong shifts (e.g. 11 % for cosmic shear+GGL). Larger fluctuations in the PJ-HPD intervals widths are also discernible, for instance for the runs with different covariance models (IDs 9–12) where the standard approach yields nearly identical constraints.

We note that, for an assessment of consistency with *Planck* constraints, the *Planck* likelihood will need to be sampled in S_8 as well. The impact of the choice of sampling parameter and the displacement of one-dimensional marginal constraints away from the underlying best-fit value highlight that discrepancies between probes should be interpreted with care if quantified solely in the marginal distributions, as is widespread practice. Tension is assessed most meaningfully in the shared parameter space of the full posterior distributions, but this can be challenging in practice. Therefore, we still assess tension via the marginal distributions of S_8, but only for probe combinations for which we find projection effects to be negligible. A detailed discussion of tension assessment is provided in Heymans et al. (2020). Bayesian approaches to performing consistency tests in large parameter spaces exist (Köhlinger et al. 2019) and have been applied to the internal consistency of KiDS-1000 cosmic shear (Asgari et al. 2020b). We advocate the use of a summary measure of tension that minimises the sensitivity to prior choices (see e.g. Handley & Lemos 2019b; Lemos et al. 2020).

We find no measurable difference between M-HPD and PJ-HPD S_8 credible intervals for *Planck* (as expected since the cosmological posterior is close to Gaussian), and small shifts for KV450 in line with the results of Fig. 22, so that the previously reported S_8 tension between these probes remains valid. The marginal S_8 posterior for KiDS-1000 cosmic shear only and for the joint analysis of clustering and weak lensing peaks close to the input and is well approximated by a Gaussian (cf. Fig. 23), so direct comparison with *Planck* is possible for this case as well.

8. Summary and conclusions

In this work we presented the methodology for a joint analysis of spectroscopic galaxy clustering from BOSS and of weak gravitational lensing from the fourth data release of the Kilo-Degree Survey (KiDS-1000). This includes a detailed investigation of the analysis choices from the galaxy catalogue level onwards to cosmological inference, taking into account the tightening of requirements due to the doubled survey area with respect to earlier KiDS cosmological analyses. We summarise here the major updates since the cosmic shear analysis by Hildebrandt et al. (2020a, KV450) and the BOSS+KV450 analysis by Tröster et al. (2020b).

Galaxy-galaxy lensing (GGL), the cross-correlation between lens galaxy positions and source galaxy ellipticities, is now incorporated into the joint analysis, using lens galaxies from both the BOSS and 2dFLenS surveys, which cover 85 % of the KiDS-1000 area. However, we found that it adds negligible amounts of cosmological constraining power in KiDS-1000 mainly because of an order of magnitude larger survey area available for clustering. Nonetheless, we identified a number of issues that will come to the fore in analyses where clustering, cosmic shear, and GGL are measured over the same footprint: our hybrid matter-galaxy power spectrum model blends perturbative and non-perturbative approaches customary in clustering and weak lensing, respectively, but a satisfactory solution for the deeply non-linear regime remains to be found (Sect. 2.2); magnification bias is strongest in the GGL signals and requires dedicated simulations if lens samples deviate from pure flux-limit selection (Appendix B); and widely used idealisations in covariance models related to survey geometry fail most prominently in the GGL part (Sect. 5.3).

Weak lensing signals are consistently modelled using band powers derived from correlation function measurements that make them insensitive to survey geometry (Sects. 2.4 and 2.5). As opposed to earlier KiDS studies, we now fully incorporate mode mixing in the modelling of both signals and covariances, but we find that on scales that we can model accurately ($\ell \in [100, 1500]$) the recovery of the underlying angular power spectra is highly accurate. The limiting systematic in the cosmological modelling is the knowledge of the highly non-linear matter power spectrum (Sect. 2.1), where current fit formulae and emulators disagree at the few per-cent level (cf. the comparison in Euclid Collaboration et al. 2019).

The main source of uncertainty in the modelling of weak lensing, however, remains the intrinsic alignment of galaxies (Sect. 2.4). Direct observations of the effect in galaxy samples typically used for weak lensing do not yet exist, and our physical understanding of alignment mechanisms is still too poor to create predictive ab-initio simulations. In the absence of clear guidance we are thus required to strike a balance between a simple model that risks not capturing the complexity in the data and a flexible model that risks providing catch-all parameters in the likelihood analysis among which other residual systematics, notably those related to the source redshift distributions, can hide.

Recent advances in selecting source galaxy samples whose redshift distribution can be more reliably calibrated (Wright et al. 2020a) have been adopted and the corresponding calibration uncertainties fully propagated into the likelihood analysis (Sect. 3.3). We also clarify how the multiplicative shear bias calibration affects estimators and derived survey characteristics critical for covariance calculation (Appendix C), and we take a more nuanced approach in quantifying and propagating the uncertainty in this calibration (Sect. 3.4).

We build a dedicated suite of more than 20 000 mocks to perform an unprecedentedly detailed assessment of our analytic covariance models (Sects. 4 and 5). KiDS-1000 cosmological constraints are demonstrated to remain unchanged when using an analytic or mock-based covariance (Sect. 7). We show that the analytic model is capable of capturing the impact of survey geometry and spatial variations in survey depth well and identify the mixed noise-sample variance contribution as the main culprit behind residual differences. We confirm recent work by Heydenreich et al. (2020) in showing that variable depth biases two-point weak lensing statistics at the few per-cent level (Sect. 4.3), which has negligible impact for KiDS-1000, but will challenge widespread analysis approaches that rely on spatial homogeneity of the galaxy sample in future applications.

We explicitly map the complex prior space of cosmic shear cosmological analyses (Sect. 6.1), which serves to explain the significant prior dependence of results (e.g. Joudaki et al. 2020). That said, the key parameter S_8 is robust to different choices of prior, especially the upper tail of its marginal posterior which directly impacts on tension measurements with *Planck*. In contrast to previous analyses, we directly sample in S_8, which allows us to impose a wide top-hat, and thus safely uninformative, prior. We employ our mock suite to demonstrate that the weak lensing band power likelihood is consistent with being Gaussian (whereas there is evidence for non-Gaussianity for correlation functions on large scales) and that the χ^2 goodness of fit sampling distribution is indeed χ^2-distributed to good approximation (Sects. 6.2 and 6.3). However, significant deviations from the textbook approach are seen for the degrees of freedom, or equivalently the effective number of model parameters, and we discuss and advocate alternative estimators.

It is demonstrated that the KiDS-1000 likelihood analysis pipeline recovers the input parameters of a mock data vector with per-cent level accuracy (Sect. 7), using a dedicated maximum a posteriori (MAP) estimate of the multivariate posterior density. Experiments have hitherto typically reported the marginal mean or mode of the S_8 posterior, which we show to be significantly shifted towards lower values from the truth, beyond 1σ for some probe combinations. This is not indicative of any biases in the inference process, but simply a consequence of a large parameter space with multiple non-linear degeneracies and complex prior volume. Tension assessment in S_8 between experiments should therefore proceed with caution if based on its marginal posterior. We develop an alternative technique that produces a credible interval from the multivariate highest posterior density (PJ-HPD) which is guaranteed to encompass the MAP estimate (Sect. 6.4).

With the fiducial KiDS-1000 analysis setup, we can expect cosmic shear-only S_8 constraints that are 20 % tighter than KV450, and joint clustering and weak lensing constraints that improve in S_8 by 29 % over the previous KiDS+BOSS analysis (Sect. 7). These changes undersell the true gain in statistical power somewhat because, as opposed to the correlation function-based analysis in KV450, S_8 does not precisely capture the direction transverse to the typical weak lensing $\Omega_m-\sigma_8$ degeneracy any more.

The key scientific question in the KiDS-1000 analysis is the tentative discrepancy in the amplitude of the matter density fluctuation power spectrum seen between low-redshift large-scale structure probes and the cosmic microwave background as observed by *Planck*. Barring the caveats with regards to the interpretation of marginal S_8 constraints and the remaining uncertainties in the modelling of galaxy astrophysics, we confirm our cosmic shear and GGL analysis pipeline to be robust and find no known systematic effect in the cosmological and astrophysical modelling or the analysis methodology that would bias S_8 by more than 0.1σ, which is also the numerical accuracy limit for parameter constraints extracted from our sampled posteriors. Together with the expected boost in constraining power, KiDS-1000 thus has the potential to give us a clear steer in the direction of statistical fluke, unknown systematic, or new physics as the most likely explanation for the tension. It should be noted that all methodology investigations presented here were performed while the real shear catalogues were still blinded.

The on-going efforts in maximising the cosmological information and fidelity for the current generation of large-scale structure surveys will feed directly into the forthcoming array of even more ambitious projects. New surveys that will start within the coming three or so years include the ESA *Euclid* mission[25] (Laureijs et al. 2011), the Vera C. Rubin Observatory LSST[26] (LSST Dark Energy Science Collaboration 2012), as well as the DESI[27] (DESI Collaboration et al. 2016) , 4MOST[28] (Richard et al. 2019), and PFS[29] (Takada et al. 2014) surveys. The combination of spectroscopic 'lens' galaxy samples with 'source' samples used for weak lensing in overlapping sky regions, as pursued in this work, will remain an important approach to exploiting these data sets. The tighter accuracy requirements in step with the vastly increased raw constraining power, the greater depth of observations, and the sheer number of galaxies observed will pose a multitude of challenges in the analysis methodology that are yet to be tackled.

Author Contributions: All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (BJ, CAL, MA, TT, CH), 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., Abdalla, F. B., Allam, S., et al. 2016, Phys. Rev. D, 94, 022001
Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526
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
Alcock, C. & Paczynski, B. 1979, Nature, 281, 358
Amara, A. & Réfrégier, A. 2007, MNRAS, 381, 1018
Asgari, M., Friswell, I., Yoon, M., et al. 2020a, MNRAS[arXiv:2004.07811]
Asgari, M., Heymans, C., Hildebrandt, H., et al. 2019, A&A, 624, A134
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2020b, arXiv e-prints, arXiv:2007.15633
Asgari, M. & Schneider, P. 2015, A&A, 578, A50
Asgari, M., Tröster, T., Heymans, C., et al. 2020c, A&A, 634, A127
Barreira, A., Krause, E., & Schmidt, F. 2018a, J. Cosmology Astropart. Phys., 2018, 053
Barreira, A., Krause, E., & Schmidt, F. 2018b, J. Cosmology Astropart. Phys., 2018, 015
Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291
Becker, M. R. & Rozo, E. 2016, MNRAS, 457, 304
Becker, M. R., Troxel, M. A., MacCrann, N., et al. 2016, Phys. Rev. D, 94, 022002
Benítez, N. 2000, ApJ, 536, 571
Bernardeau, F., van Waerbeke, L., & Mellier, Y. 1997, A&A, 322, 1
Bernstein, G. M. 2009, ApJ, 695, 652
Blake, C., Amon, A., Asgari, M., et al. 2020, A&A, 642, A158
Blake, C., Amon, A., Childress, M., et al. 2016, MNRAS, 462, 4240
Blake, C., Brough, S., Colless, M., et al. 2011, MNRAS, 415, 2876
Blazek, J. A., MacCrann, N., Troxel, M. A., & Fang, X. 2019, Phys. Rev. D, 100, 103506
Bose, B., Winther, H. A., Pourtsidou, A., et al. 2020, J. Cosmology Astropart. Phys., 2020, 001
Bridle, S. & King, L. 2007, New Journal of Physics, 9, 444
Cacciato, M., Lahav, O., van den Bosch, F. C., Hoekstra, H., & Dekel, A. 2012, MNRAS, 426, 566
Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, MNRAS, 447, 646
Carron, J. 2013, A&A, 551, A88
Chan, K. C., Scoccimarro, R., & Sheth, R. K. 2012, Phys. Rev. D, 85, 083509
Chisari, N. E., Mead, A. J., Joudaki, S., et al. 2019, The Open Journal of Astrophysics, 2, 4
Colas, T., d'Amico, G., Senatore, L., Zhang, P., & Beutler, F. 2020, J. Cosmology Astropart. Phys., 2020, 001

Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, MNRAS, 453, 1513

Crocce, M. & Scoccimarro, R. 2006, Phys. Rev. D, 73, 063519

Crocce, M., Scoccimarro, R., & Bernardeau, F. 2012, MNRAS, 427, 2537

d'Amico, G., Gleyzes, J., Kokron, N., et al. 2020, J. Cosmology Astropart. Phys., 2020, 005

de Putter, R., Wagner, C., Mena, O., Verde, L., & Percival, W. J. 2012, J. Cosmology Astropart. Phys., 2012, 019

Dekel, A. & Lahav, O. 1999, ApJ, 520, 24

Deshpande, A. C., Kitching, T. D., Cardone, V. F., et al. 2020, A&A, 636, A95

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

Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64

Duncan, C. A. J., Joachimi, B., Heavens, A. F., Heymans, C., & Hildebrandt, H. 2014, MNRAS, 437, 2471

Dvornik, A., Hoekstra, H., Kuijken, K., et al. 2018, MNRAS, 479, 1240

Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32

Efstathiou, G. & Lemos, P. 2018, MNRAS, 476, 151

Eggemeier, A., Scoccimarro, R., Crocce, M., Pezzotta, A., & Sánchez, A. G. 2020, Phys. Rev. D, 102, 103530

Eifler, T., Krause, E., Dodelson, S., et al. 2015, MNRAS, 454, 2451

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

Erben, T., Schirmer, M., Dietrich, J. P., et al. 2005, Astronomische Nachrichten, 326, 432

Euclid Collaboration, Knabenhans, M., Stadel, J., et al. 2019, MNRAS, 484, 5509

Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23

Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, MNRAS, 467, 1627

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, MNRAS[arXiv:2003.02700]

Fosalba, P., Crocce, M., Gaztañaga, E., & Castand er, F. J. 2015a, MNRAS, 448, 2987

Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, MNRAS, 447, 1319

Gelman, A., Meng, X.-L., & Stern, H. 1992, Tech. Report #355, Dept. of Statistics, University of Chicago

Gelman, A., Meng, X.-L., & Stern, H. 1996, Statistica Sinica, 6, 733

Giblin, B., Heymans, C., Asgari, M., et al. 2020, arXiv e-prints, arXiv:2007.01845

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Griffith, R. L., Cooper, M. C., Newman, J. A., et al. 2012, ApJS, 200, 9

Hamilton, A. J. S., Kumar, P., Lu, E., & Matthews, A. 1991, ApJ, 374, L1

Hamimeche, S. & Lewis, A. 2008, Phys. Rev. D, 77, 103013

Handley, W. & Lemos, P. 2019a, Phys. Rev. D, 100, 023512

Handley, W. & Lemos, P. 2019b, Phys. Rev. D, 100, 043504

Harnois-Déraps, J., Amon, A., Choi, A., et al. 2018, MNRAS, 481, 1337

Harnois-Déraps, J., Giblin, B., & Joachimi, B. 2019, A&A, 631, A160

Hartlap, J., Hilbert, S., Schneider, P., & Hildebrand t, H. 2011, A&A, 528, A51

Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399

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

Heymans, C., Tröster, T., Asgari, M., et al. 2020, arXiv e-prints, arXiv:2007.15632

Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146

Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43

Hilbert, S., Hartlap, J., & Schneider, P. 2011, A&A, 536, A85

Hildebrandt, H. 2016, MNRAS, 455, 3943

Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020a, A&A, 633, A69

Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2020b, arXiv e-prints, arXiv:2007.15635

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454

Hirata, C. M. & Seljak, U. 2004, Phys. Rev. D, 70, 063526

Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, MNRAS, 447, 1724

Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, J. Cosmology Astropart. Phys., 2012, 027

Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592

Huang, H.-J., Eifler, T., Mandelbaum, R., & Dodelson, S. 2019, MNRAS, 488, 1652

Huff, E. & Mandelbaum, R. 2017, arXiv e-prints, arXiv:1702.02600

Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, J. Cosmology Astropart. Phys., 2020, 042

Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338

Joachimi, B. & Bridle, S. L. 2010, A&A, 523, A1

Joachimi, B., Mandelbaum, R., Abdalla, F. B., & Bridle, S. L. 2011, A&A, 527, A26

Joachimi, B., Schneider, P., & Eifler, T. 2008, A&A, 477, 43

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

Kaiser, N. 1992, ApJ, 388, 272

Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92

Kaufman, G. M. 1967, Some Bayesian Moment Formulae, Report No. 6710, Center for Operations Research and Econometrics, Catholic University of Louvain, Heverlee, Belgium

Kayo, I., Taruya, A., & Suto, Y. 2001, ApJ, 561, 22

Kazin, E. A., Sánchez, A. G., & Blanton, M. R. 2012, MNRAS, 419, 3223

Kilbinger, M. & Schneider, P. 2004, A&A, 413, 465

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

Köhlinger, F., Viola, M., Joachimi, B., et al. 2017, MNRAS, 471, 4412

Kohonen, T. 1982, Biological Cybernetics, 43, 59

Kosowsky, A., Milosavljevic, M., & Jimenez, R. 2002, Phys. Rev. D, 66, 063007

Krause, E. & Eifler, T. 2017, MNRAS, 470, 2100

Krause, E., Eifler, T. F., Zuntz, J., et al. 2017, arXiv e-prints, arXiv:1706.09359

Krause, E. & Hirata, C. M. 2010, A&A, 523, A28

Kuijken, K. 2011, The Messenger, 146, 8

Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2

Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500

Kunz, M., Trotta, R., & Parkinson, D. R. 2006, Phys. Rev. D, 74, 023503

Landy, S. D. & Szalay, A. S. 1993, ApJ, 412, 64

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193

Lemos, P., Köhlinger, F., Handley, W., et al. 2020, MNRAS, 496, 4647

Lewis, A. & Bridle, S. 2002, Phys. Rev. D, 66, 103511

Lewis, A., Challinor, A., & Lasenby, A. 2000, The Astrophysical Journal, 538, 473

Li, Y., Hu, W., & Takada, M. 2014, Phys. Rev. D, 89, 083519

Lin, C.-H., Harnois-Déraps, J., Eifler, T., et al. 2020, MNRAS, 499, 2977

Loverde, M. & Afshordi, N. 2008, Phys. Rev. D, 78, 123506

LSST Dark Energy Science Collaboration. 2012, arXiv e-prints, arXiv:1211.0310

Mandelbaum, R. 2018, ARA&A, 56, 393

Mandelbaum, R., Hirata, C. M., Seljak, U., et al. 2005, MNRAS, 361, 1287

Masters, K. L., Maraston, C., Nichol, R. C., et al. 2011, MNRAS, 418, 1055

McFarland, J. P., Verdoes-Kleijn, G., Sikkema, G., et al. 2013, Experimental Astronomy, 35, 45

Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, MNRAS, 454, 1958

Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858

Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

Park, Y., Rozo, E., & Krause, E. 2020, arXiv e-prints, arXiv:2004.07504

Percival, W. J., Ross, A. J., Sánchez, A. G., et al. 2014, MNRAS, 439, 2531

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A6

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020b, A&A, 641, A5

Price, M. A., McEwen, J. D., Pratley, L., & Kitching, T. D. 2020, MNRAS[arXiv:2004.07855]

Raveri, M. & Hu, W. 2019, Phys. Rev. D, 99, 043506

Reid, B., Ho, S., Padmanabhan, N., et al. 2016, MNRAS, 455, 1553

Richard, J., Kneib, J. P., Blake, C., et al. 2019, The Messenger, 175, 50

Robert, C. 2001, The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, Springer Texts in Statistics (Springer New York)

Ross, A. J., Beutler, F., Chuang, C.-H., et al. 2017, MNRAS, 464, 1168

Rozo, E., Rykoff, E. S., Abate, A., et al. 2016, MNRAS, 461, 1431

Salazar-Albornoz, S., Sánchez, A. G., Grieb, J. N., et al. 2017, MNRAS, 468, 2938

Samuroff, S., Blazek, J., Troxel, M. A., et al. 2019, MNRAS, 489, 5453

Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2018, MNRAS, 475, 4524

Sánchez, A. G., Scoccimarro, R., Crocce, M., et al. 2017, MNRAS, 464, 1640

Schmidt, F., Rozo, E., Dodelson, S., Hui, L., & Sheldon, E. 2009, ApJ, 702, 593

Schneider, A., Stoira, N., Refregier, A., et al. 2020, J. Cosmology Astropart. Phys., 2020, 019

Schneider, P. & Hartlap, J. 2009, A&A, 504, 705

Schneider, P., van Waerbeke, L., Jain, B., & Kruse, G. 1998, MNRAS, 296, 873

Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002a, A&A, 396, 1

Schneider, P., van Waerbeke, L., & Mellier, Y. 2002b, A&A, 389, 729

Scoccimarro, R. 2004, Phys. Rev. D, 70, 083007

Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20

Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101

Sellentin, E. & Heavens, A. F. 2018, MNRAS, 473, 2355
Sellentin, E., Heymans, C., & Harnois-Déraps, J. 2018, MNRAS, 477, 4879
Shanks, T., Metcalfe, N., Chehade, B., et al. 2015, MNRAS, 451, 4238
Sheldon, E. S. & Huff, E. M. 2017, ApJ, 841, 24
Shirasaki, M. & Takada, M. 2018, MNRAS, 478, 4277
Silverman, B. W. 1986, Density estimation for statistics and data analysis (Chapman and Hall)
Simon, P. & Hilbert, S. 2018, A&A, 613, A15
Singh, S., Mandelbaum, R., Seljak, U., Slosar, A., & Vazquez Gonzalez, J. 2017, MNRAS, 471, 3827
Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. 2002, J. R. Stat. Soc. B, 64, 583
Stölzner, B., Joachimi, B., Korn, A., Hildebrandt, H., & Wright, A. H. 2020, arXiv e-prints, arXiv:2012.07707
Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1
Takada, M. & Hu, W. 2013, Phys. Rev. D, 87, 123504
Takada, M. & Jain, B. 2004, MNRAS, 348, 897
Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152
Taruya, A., Bernardeau, F., Nishimichi, T., & Codis, S. 2012, Phys. Rev. D, 86, 103528
Thiele, L., Duncan, C. A. J., & Alonso, D. 2020, MNRAS, 491, 1746
Thomas, D. B., Whittaker, L., Camera, S., & Brown, M. L. 2017, MNRAS, 470, 3131
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878
Tröster, T., Asgari, M., Blake, C., et al. 2020a, arXiv e-prints, arXiv:2010.16416
Tröster, T., Sánchez, A. G., Asgari, M., et al. 2020b, A&A, 633, L10
Troxel, M. A., Krause, E., Chang, C., et al. 2018a, MNRAS, 479, 4998
Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018b, Phys. Rev. D, 98, 043528
Unruh, S., Schneider, P., Hilbert, S., et al. 2020, A&A, 638, A96
Vale, C., Hoekstra, H., van Waerbeke, L., & White, M. 2004, ApJ, 613, L1
van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, A&A, 642, A200
van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, MNRAS, 476, 4662
Verde, L., Treu, T., & Riess, A. G. 2019, Nature Astronomy, 3, 891
White, M., Blanton, M., Bolton, A., et al. 2011, ApJ, 728, 126
Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, A&A, 632, A34
Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020a, A&A, 637, A100
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020b, A&A, 640, L14
Xavier, H. S., Abdalla, F. B., & Joachimi, B. 2016, MNRAS, 459, 3693
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45
Zuntz, J., Sheldon, E., Samuroff, S., et al. 2018, MNRAS, 481, 1149

Appendix A: Additional tables related to signal modelling

In Table A.1 the fiducial parameter values for our analyses are shown. Unless stated otherwise, these are adopted throughout. They generally follow the rounded best-fit values from the Tröster et al. (2020b) analysis of BOSS and KV450. For weakly constrained parameters we instead choose the centre values of the flat priors adopted to avoid undue skewness in the prior volume. Magnification parameters are determined as described in Appendix B. In addition to the parameters defined in the main body of the paper, we list the curvature density parameter Ω_K, the baryon density parameter Ω_b, and the dark energy density parameter Ω_Λ. Table A.2 shows the best-fit coefficients for the galaxy-matter power spectrum fit function (Eq. 5) introduced in Sect. 2.2.

Table A.1. Choice of fiducial model parameters.

Parameter	Symbol	Value
CDM density	ω_c	0.13
Baryon density	ω_b	0.0225
Scalar spectral amp.	$\ln(10^{10}A_s)$	2.72
Scalar spectral index	n_s	0.97
Hubble constant (scaled)	h	0.7
Neutrino mass sum	$\sum m_\nu$	0.06 eV
Curvature density parameter	Ω_K	0
Matter density parameter	Ω_m	0.31
Baryon density parameter	Ω_b	0.046
Dark energy density parameter	Ω_Λ	0.69
Density fluctuation amp.	σ_8	0.733
Weak lensing amp. parameter	S_8	0.746
Linear galaxy bias	b_1	{2.1, 2.3}
Quadratic galaxy bias	b_2	{0.2, 0.5}
Non-local galaxy bias	γ_3^-	{0.9, 0.1}
Virial velocity parameter	$a_{\rm vir}$	{3.8, 3.0}
Luminosity function slope	$\alpha_{\rm mag}$	{1.8, 2.6}
Intrinsic alignment (IA) amp.	$A_{\rm IA}$	0.8
IA redshift dependence	$\eta_{\rm IA}$	0
Baryon feedback amp.	$A_{\rm bary}$	2.6
Redshift offset	δ_z	{0, 0, 0, 0, 0}

Notes. The first section contains the primordial ΛCDM parameters, the second a selection of derived cosmological parameters. The third section lists astrophysical model parameters, with pairs of fiducial values for the two lens bins L1 and L2 in curly brackets. The fourth section comprises the parameters varied in the inference related to measurement systematics (one per source galaxy sample).

Table A.2. Fit coefficients $g_{\alpha,i}^{mn}$ of the approximate galaxy-matter power spectrum model; cf. Eq. (5).

α	i	polynomial order $m\,n$					
		0 0	0 1	1 0	0 2	1 1	2 0
b_2	2	0.56	−0.96	−2.55	0.34	2.08	2.05
	1	5.03	−2.71	−5.52	0.89	5.78	3.90
	0	−0.45	−2.21	−5.69	0.24	1.49	4.12
γ_2	2	−0.03	−0.01	0.88	0.02	0.09	−1.03
	1	3.45	−0.34	1.69	0.05	−0.19	−2.47
	0	0.26	−1.86	−3.68	0.51	1.50	2.39
γ_3^-	2	−0.01	0.00	0.81	0.01	0.11	−1.01
	1	3.43	−0.32	1.17	0.01	−0.06	−2.06
	0	−1.81	−1.56	−3.08	0.49	1.50	1.93

Appendix B: Magnification bias

Gravitational lensing not only modifies the ellipticities of galaxy images but also their apparent size and, therefore, their measured flux. Since the definition of galaxy samples almost universally involves flux-dependent quantities, the selection function acquires a dependency on the foreground large-scale structure and thus on cosmology. This magnification bias is a second-order effect for gravitational shear measurements in our source samples which may need to be accounted for in future weak lensing surveys but which we can safely neglect in KiDS (Schmidt et al. 2009; Krause & Hirata 2010; Deshpande et al. 2020). However, it is not a priori clear that magnification bias is negligible in our lens samples, especially since it directly modulates the observable, that is the number counts of galaxies (see e.g. Duncan et al. 2014; Unruh et al. 2020; Thiele et al. 2020 for studies of the impact of magnification bias).

If the samples are purely flux-limited, magnification bias is readily modelled as a balance of the local dilution/focussing of solid angle and the modification of apparent magnitudes above or below the flux limit (Bartelmann & Schneider 2001), resulting in expressions of the form given in Eq. (29). The only additional information required is the logarithmic slope of the sample's luminosity function at its faint end, $\alpha_{\rm mag}$. However, like other lens samples used for large-scale structure inference jointly with weak lensing (e.g. Rozo et al. 2016), the two BOSS samples employed in the KiDS-1000 analysis were derived from a complex multi-band selection function, making closed-form analytic modelling impossible (cf. Hildebrandt 2016).

In a separate publication (von Wietersheim et al., in prep.) we develop a simulation-assisted method to determine an effective luminosity function slope for our BOSS samples that can be used in the standard expressions. Here, we summarise the salient points (for details see van den Busch et al. 2020): we recreate the LOWZ and CMASS selections in the MICE2 galaxy mocks (Fosalba et al. 2015a,b; Crocce et al. 2015) over an octant of the sky. Two variants of the BOSS DR12 low- and high-redshift samples are then derived, one with the selection applied to observed fluxes that include magnification, and one with the selection based on the hypothetical fluxes if no magnification was present. From this we can directly measure $\alpha_{\rm mag}$ via the slope of the linearised relation between the change in galaxy number counts between the magnified and unmagnified catalogues and the weak lensing convergence of these objects. The result is compared to the slope of the magnitude count of the same mock samples near where the counts begin to drop off, and we identify a magnitude range within which the slope measurements agree within the noise.

We find that for both the low- and high-redshift samples i-band counts yield the best match, driven by the i-band cut of the original CMASS sample which also substantially contributes to L1. The latter measurement is repeated on the real BOSS catalogues, using the same band and magnitude range as identified in the mock. This results in

$$\alpha_{\rm mag}^{\rm L1} = 1.80 \pm 0.15 \,; \qquad \alpha_{\rm mag}^{\rm L2} = 2.62 \pm 0.28 \,, \tag{B.1}$$

where the errors are the standard deviations of the magnitude counts among the bins selected for the slope measurements. Rather than pure statistical noise, these errors therefore have contributions from the limited accuracy of the assumption that the BOSS-selected samples have a simple power-law scaling in the magnitude range that we consider.

We propagate these $\alpha_{\rm mag}$ measurements into angular power spectrum predictions for clustering and galaxy-galaxy sig-
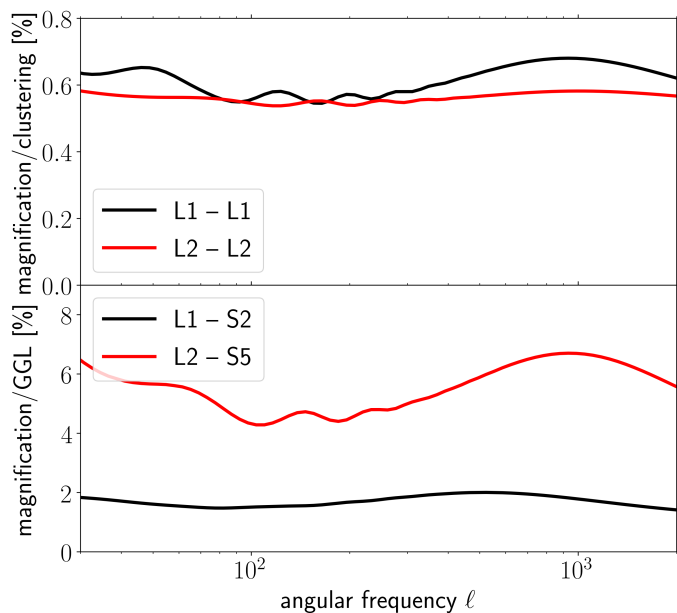
Fig. B.1. Magnification bias contribution relative to galaxy clustering (top panel) and the galaxy-galaxy lensing signal (bottom panel) for the redshift bin combinations indicated in the legends (cf. Table 2). Shown are angular power spectrum models assuming linear galaxy bias using the values from Table A.1.

nals; see Fig. B.1. In both lens bins the magnification contributions (dominated by the magnification-clustering cross-correlation) are at the 0.6 % level and therefore negligible. The magnification-shear correlation constitutes a few-per cent contribution to the galaxy-galaxy lensing signal and is larger for the high-redshift source bins. This result prompts us to include the latter contributions into our modelling, as further described in Sect. 2.5, but they are too small to justify additional freedom in the GGL model, so that we keep the α_{mag} parameters fixed at their best-fit values (Eq. B.1) in the analysis.

Appendix C: Propagation of multiplicative shear calibration

In this appendix we show how the multiplicative shear calibration propagates into the correlation function estimators, the sample redshift distributions, and the sample properties that are inputs for the analytic covariance calculation. We work in the weak shear limit where $\gamma \ll 1$, and will model our shear estimator, the observed ellipticity ϵ^{obs}, as

$$\epsilon^{\mathrm{obs}} = (1 + m)[\epsilon + \gamma] , \tag{C.1}$$

working in the absence of additive biases such as PSF contamination. Here the unsheared ellipticity ϵ is uncorrelated with γ (i.e. there are no intrinsic alignments), and is a combination of the true intrinsic ellipticity ϵ^{int} and random measurement noise. As noise increases in the imaging, galaxy shapes become increasingly round and our ability to recover the underlying shear decreases, independently of the shear estimation method used. We express this inherent fundamental noise bias in terms of a multiplicative shear calibration correction m that can be seen as both a correction that provides an unbiased shear estimator, and also a weight, reflecting that some galaxies carry little shear information. Regarding this latter point, using the term 'Responsivity' or 'Sensitivity' to describe $(1 + m)$ is useful (see e.g. Huff

& Mandelbaum 2017; Sheldon & Huff 2017; Zuntz et al. 2018) as this term quantifies how responsive, or sensitive, the observed galaxy is to an induced shear. Unfortunately we cannot measure m precisely for each galaxy and can only determine an estimate $\hat{m}$, where

$$\hat{m} = m + \eta . \tag{C.2}$$

Here, η is a noise term that has zero mean. It can potentially have a very significant variance such that one should not attempt to de-bias ϵ^{obs} individually for each galaxy because of the non-linear propagation of the noise in $\hat{m}$.

We note in passing that our multiplicative shear bias calibration takes the impact of variable survey depth on shear measurement into account because $\hat{m}$ is a function of the signal-to-noise and size of the galaxy image, simulated for a range of different seeings (see Giblin et al. 2020 for details). While we do not take into account that in a poor-seeing, low-depth region the fraction of blended galaxies will be higher than average, Kannawadi et al. (2019) showed the calibration to be robust to changes in galaxy background density and thus the levels of blending well within the derived calibration uncertainty. We therefore do not expect the coupling of blending-induced shear bias with variable depth to make a significant impact on the parameter constraints, as artificially changing the blend fraction across the whole simulated survey changes $\hat{m}$ by an insignificant amount.

Appendix C.1: Shear correlation function estimator

The typical starting point for measuring cosmic shear are the two-point correlation function estimators $\hat{\xi}_{\pm}$ that we could ideally write as

$$\hat{\xi}_{\pm}(\theta) = \frac{\sum_{ij} w_i w_j (\epsilon_{it}^{\mathrm{obs}} \epsilon_{jt}^{\mathrm{obs}} \pm \epsilon_{i\times}^{\mathrm{obs}} \epsilon_{j\times}^{\mathrm{obs}}) \Delta_{ij}(\theta)}{\sum_{ij} w_i w_j (1 + \hat{m}_i)(1 + \hat{m}_j) \Delta_{ij}(\theta)} , \tag{C.3}$$

where w_i is a survey-defined weight that has been assigned to galaxy i, and the sum is taken over all galaxies i and j. We have also introduced the selector function $\Delta_{ij}(\theta)$, which is unity if the angular separation between galaxies i and j lies within a bin centred on θ, and zero otherwise. The tangential and cross components of the ellipticity (and analogously for the shear) are given by $\epsilon_{\mathrm{t}} + i\epsilon_{\times} = -\epsilon \, e^{-2i\varphi}$, where φ is the polar angle of the separation vector between the galaxy pair under consideration.

Ignoring contributions from intrinsic galaxy alignments, we can expand this estimator using Eq. (C.1) as

$$\hat{\xi}_{\pm}(\theta) = \frac{\sum_{ij} W_i W_j (\epsilon_{it} \epsilon_{jt} \pm \epsilon_{i\times} \epsilon_{j\times}) \Delta_{ij}(\theta)}{\sum_{ij} \widehat{W}_i \widehat{W}_j \Delta_{ij}(\theta)}$$
$$+ \frac{\sum_{ij} W_i W_j (\gamma_{it} \gamma_{jt} \pm \gamma_{i\times} \gamma_{j\times}) \Delta_{ij}(\theta)}{\sum_{ij} \widehat{W}_i \widehat{W}_j \Delta_{ij}(\theta)} , \tag{C.4}$$

where $W_i := w_i(1 + m_i)$ and $\widehat{W}_i := w_i(1 + \hat{m}_i)$. The first term introduces shape noise into our analysis (which we explore in Sect. C.3). The second term is the true weighted cosmic shear signal that we wish to extract, where we can see that our weighted source galaxy distribution now includes the Responsivity, correctly down-weighting galaxies that are unresponsive to shear (as these typically have negative values for m).

In practice, Eq. C.3 is not the actual estimator that we employ, as we assume that the weights and responses are uncorrelated, which is a good approximation, such that the denominator

becomes

$$K(\theta) := \sum_{i \in I, j \in J} \widehat{W}_i \widehat{W}_j \, \Delta_{ij}(\theta) \tag{C.5}$$

$$\approx (1 + \langle m \rangle_I)(1 + \langle m \rangle_J) \sum_{i \in I, j \in J} w_i \, w_j \, \Delta_{ij}(\theta),$$

where we have made the samples over which the summations run explicit, and where $\langle m \rangle_I$ denotes the average of $\hat{m}$ over all galaxies in sample I. In taking this sum we reduce the impact of noise in our calibration correction and recover the average noise-bias for the population. This correction can however lead to the misconception that $K(\theta)$ is only to calibrate the average value of the shear for the galaxy sample. It is, however, also taking into account the effective down-weighting of unresponsive galaxies that is an inherent part of the shear estimator ϵ^{obs}.

Appendix C.2: Redshift distribution

From Eq. C.4 we can see that source galaxies used in the cosmic shear measurement are weighted by W, which is a combination of the survey-defined weight w and the shear sensitivity $1 + m$. An estimate of the effective redshift distribution of this source sample is therefore given by

$$n_{\mathrm{S}}(z) = \frac{\sum_i \widehat{W}_i \, n_i(z)}{\sum_i \widehat{W}_i} = \frac{\sum_i w_i(1 + m_i + \eta_i)\, n_i(z)}{\sum_i w_i(1 + m_i + \eta_i)}, \tag{C.6}$$

where $n_i(z)$ is the probability distribution of true redshifts for an individual galaxy i. We note that in Eq. (C.6) we have made the common approximation that the redshift PDF of a union of galaxy sub-samples is the weighted sum of the sub-sample PDFs. We use $\widehat{W}$ in Eq. (C.6) as only the noisy weight is accessible from observations. However, as the m-noise η has zero mean, this term does not impact the redshift estimation. If m is uncorrelated with redshift z then this term will cancel in the estimate. As m is correlated with size and magnitude, however, it is likely to correlate with z and therefore it is necessary to include it as part of the weight in the effective redshift distribution.

For the DES Year 1 analysis, Eq. (C.6) corresponds to their weighted stack of individual galaxy PDFs estimated using BPZ (Hoyle et al. 2018; see also Sheldon & Huff 2017). In KiDS-1000, we determine the true redshift z_μ per cell μ in the self-organising map (SOM; see Sect. 3.3) and construct an effective redshift distribution for the full source sample as

$$n_{\mathrm{S}}(z) = \frac{\sum_\mu n_\mu(z) \sum_{i \in \mu} \widehat{W}_i}{\sum_i \widehat{W}_i}, \tag{C.7}$$

where the sum $\sum_{i \in \mu}$ runs over all galaxies i in SOM-cell μ, and $n_\mu(z) = \delta_{\mathrm{D}}(z - z_\mu)$, where δ_{D} denotes the Dirac delta-distribution. We construct a multiplicative shear calibration estimate m per galaxy by applying 'Method A' from Fenech Conti et al. (2017) to the image simulations of Kannawadi et al. (2019). This involves fitting m as a function of signal-to-noise and size (see for example Fig. 9 in Fenech Conti et al. 2017). Fenech Conti et al. (2017) find the accuracy of these m-per galaxy estimates to be lower than the m-per sample estimates that we adopt in our fiducial analysis. They are, however, sufficient to determine the impact of including the Responsivity in our redshift estimates. We find the difference to be negligible between the redshift distribution calculated with Eq. (C.7) when incorporating the Responsivity, or when setting $m = 0$. This is because m is typically small for the self-calibrating *lens*fit approach that KiDS takes, but this need not be the case, such as for metacalibration approaches.

Appendix C.3: Shape noise estimates

Shape noise, quantified via σ_ϵ, is often defined in the literature as 'the intrinsic ellipticity dispersion' and is a crucial ingredient into the analytical covariance calculation. The definition needs to be reconsidered for the case of a weighted and calibrated ellipticity distribution. Schneider et al. (2002a) derive the analytical covariance for $\xi_\pm$. In their Equation (13), σ_ϵ is defined as

$$\langle \epsilon_{it}^{\mathrm{obs}} \epsilon_{jt}^{\mathrm{obs}} + \epsilon_{i\times}^{\mathrm{obs}} \epsilon_{j\times}^{\mathrm{obs}} \rangle = \sigma_\epsilon^2 \, \delta_{ij} + \xi_+(|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j|), \tag{C.8}$$

where $\boldsymbol{\theta}_i$ denotes the angular position vector of galaxy i. This means that the effective σ_ϵ for our shear correlation function estimator is equal to $\hat{\xi}_+(\theta = 0)$ in a universe with vanishing gravitational shear (for zero lag, $\delta_{ij} = 1$ in Eq. C.8). Applying this condition to Eq. (C.3), the shape noise can therefore be estimated from our weighted and calibrated ellipticity distribution as

$$\sigma_\epsilon^2 = K^{-1}(0) \sum_i w_i^2 \left[(\epsilon_{i1}^{\mathrm{obs}})^2 + (\epsilon_{i2}^{\mathrm{obs}})^2 \right], \tag{C.9}$$

where we have approximated the intrinsic ellipticity dispersion by the observed ellipticity dispersion as $|\gamma| \ll 1$, that is $W\epsilon \approx w\epsilon^{\mathrm{obs}}$. In evaluating $K(0)$ we use the limit $\Delta_{ij}(0) = \delta_{ij}$. The variances of the ellipticities are calculated after the additive bias correction has been applied (as discussed in Giblin et al. 2020).

Appendix C.4: Effective galaxy pair count and number density

Since sample variance contributions are independent of how densely the shear field is sampled, we can restrict ourselves to the shape noise contribution to the correlation function covariance. This term is readily isolated by once again considering the case of vanishing gravitational shear. In this limit the covariances of ξ_+ and ξ_- are identical and only have a diagonal contribution. Schneider et al. (2002a) derived a general expression for this covariance term, which we can adapt to our estimator (C.3) as follows:

$$\mathrm{Cov}_{\mathrm{G,sn}} [\xi_+(\theta); \xi_+(\theta)] = \mathrm{Cov}_{\mathrm{G,sn}} [\xi_-(\theta); \xi_-(\theta)]$$

$$= K^{-2}(\theta) \sum_{ijkl} W_i W_j W_k W_l \, \Delta_{ij}(\theta) \, \Delta_{kl}(\theta)$$

$$\times \langle \epsilon_{it} \epsilon_{jt} \epsilon_{kt} \epsilon_{lt} + \epsilon_{it} \epsilon_{jt} \epsilon_{k\times} \epsilon_{l\times} + \epsilon_{i\times} \epsilon_{j\times} \epsilon_{kt} \epsilon_{lt} + \epsilon_{i\times} \epsilon_{j\times} \epsilon_{k\times} \epsilon_{l\times} \rangle$$

$$= \frac{\sigma_\epsilon^4}{K^2(\theta)} \sum_{ij} W_i^2 W_j^2 \, \Delta_{ij}(\theta). \tag{C.10}$$

We note that off-diagonal terms correlating $\xi_\pm$ at different angular separations vanish, and that $\mathrm{Cov}_{\mathrm{G,sn}} [\xi_+(\theta); \xi_-(\theta)] = 0$. In the absence of weighting and calibration this expression reduces to the intuitive form $\mathrm{Cov}_{\mathrm{G,sn}} [\xi_\pm(\theta); \xi_\pm(\theta)] = \sigma_\epsilon^4/N_{\mathrm{pair}}(\theta)$, where $N_{\mathrm{pair}}(\theta)$ is the number of pairs among the galaxy samples correlated and within the survey footprint with a separation that falls into the bin centred on θ. Retaining this expression, we can define an effective number of galaxy pairs in the presence of weights and calibration as

$$N_{\mathrm{pair}}(\theta) := \frac{K^2(\theta)}{\sum_{ij} W_i^2 W_j^2 \, \Delta_{ij}(\theta)} \approx \frac{K^2(\theta)}{\sum_{ij} \widehat{W}_i^2 \widehat{W}_j^2 \, \Delta_{ij}(\theta)}, \tag{C.11}$$

such that the calibration correction carries through correctly into the covariance matrix when we measure this effective number of pairs directly from the data. While the first equality in Eq. (C.11)

is exact, the second one replaces the unobservable noiseless W with the observable quantity $\widehat{W}$ and is in practice used on the data. The effective number density then follows by setting $\Delta_{ij} \equiv 1$, that is by calculating all galaxy pairs in the survey, and using the relation $N_{\text{pair}}^{\text{total}} = A_{\text{eff}}^2 n_{\text{eff,S}}^2$ to obtain:

$$n_{\text{eff,S}} = \frac{1}{A_{\text{eff}}} \frac{(\sum_i \widehat{W}_i)^2}{\sum_i W_i^2} \approx \frac{1}{A_{\text{eff}}} \frac{(\sum_i \widehat{W}_i)^2}{\sum_i \widehat{W}_i^2}, \quad (C.12)$$

where A_{eff} is the effective survey area further discussed in Appendix E. Again, the second equality is an approximative expression that is applied to the data. The expressions above generalise to the tomographic case in a straightforward manner.

The results for σ_ϵ, N_{pair}, and n_{eff} derived here reduce to the expressions used in previous analyses (Heymans et al. 2012; Kuijken et al. 2015) for unit Responsivity. The changes in these quantities due to multiplicative shear calibration are small in our analysis, again because m is typically small for the self-calibrating *lens*fit approach that KiDS takes.

Appendix C.5: Tangential shear case

We repeat the steps above for the case of galaxy-galaxy lensing, using the tangential shear estimator of Eq. (38). Requiring that the noise term in the diagonal elements of the covariance of this estimator is rendered as $\text{Cov}_{\text{G,sn}}[\langle\gamma_t\rangle(\theta); \langle\gamma_t\rangle(\theta)] = \sigma_\epsilon^2/(2N_{\text{pair}}^{\text{GGL}})$, one obtains

$$N_{\text{pair}}^{\text{GGL}}(\theta) := \frac{\left(\sum_{i\in R, j\in S} w_i \widehat{W}_j \Delta_{ij}(\theta)\right)^2}{\mathcal{N}_{\text{rnd}}^2 \left(\sum_{i\in L, j\in S} w_i^2 W_j^2 \Delta_{ij}(\theta)\right)}, \quad (C.13)$$

where S, L, and R denote the source, lens, and random catalogue, respectively. We again set $\Delta_{ij} \equiv 1$ and identify $n_{\text{eff,S}}$ as given by Eq. (C.12) in the resulting expression. Assuming $N_{\text{pair}}^{\text{GGL,total}} = A_{\text{eff}}^2 n_{\text{eff,S}} n_{\text{eff,L}}$, we arrive at the following equation for the effective number density of lens galaxies:

$$n_{\text{eff,L}} = \frac{1}{A_{\text{eff}}} \frac{(\sum_{i\in R} w_i)^2}{\mathcal{N}_{\text{rnd}}^2 \left(\sum_{i\in L} w_i^2\right)} = \frac{1}{A_{\text{eff}}} \frac{(\sum_{i\in L} w_i)^2}{\left(\sum_{i\in L} w_i^2\right)}, \quad (C.14)$$

so the expression is fully analogous to the one for $n_{\text{eff,S}}$.

Appendix D: Analysis of covariance models

This section covers additional topics in covariance modelling; see Sect. 5 for an overview. The construction of the analytic model employed here is detailed in Appendix E.

Appendix D.1: Clustering - weak lensing cross-variance

To demonstrate that the clustering and weak lensing signals are uncorrelated and thus statistically independent (in the Gaussian likelihood approximation), we create 4000 mock realisations with the full BOSS footprint, KiDS, and their overlap. We do not simulate 2dFLenS lens galaxies as the additional GGL measurement outside the area from which clustering was obtained only acts to further reduce correlations. The current FLASK implementation does not allow us to incorporate line-of-sight modes so that we cannot model the redshift-space correlation function. However, since weak lensing only depends on transverse modes
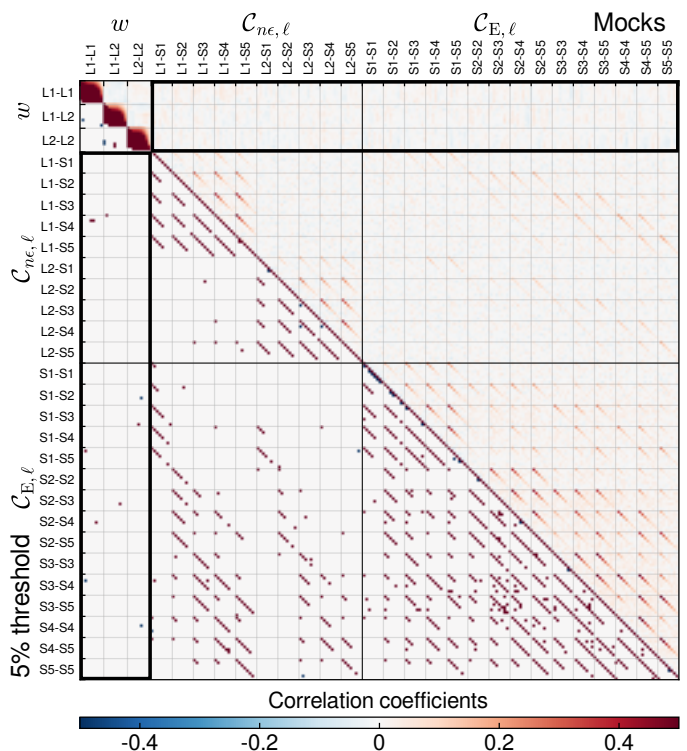
Fig. D.1. Correlation coefficient matrix for the angular galaxy clustering correlation function w, the galaxy-galaxy lensing band power $C_{n\epsilon}$, and the cosmic shear band power C_E calculated over the full BOSS and KiDS-1000 footprints. The upper triangle shows the correlation matrix as calculated from the mocks; the lower triangle indicates correlation coefficients above $\pm 5\%$. The black rectangles highlight the cross-correlations between galaxy clustering and weak lensing statistics. These are negligible, with few elements fluctuating above the 5% threshold.

of the density distribution, it is sufficient to consider the projected angular correlation function,

$$w^{(ij)}(\theta) = \int_0^\infty \frac{\mathrm{d}\ell\, \ell}{2\pi} J_0(\ell\theta)\, C_{\text{gg}}^{(ij)}(\ell), \quad (D.1)$$

where C_{gg} is the angular galaxy power spectrum. The correlation function is measured with the standard Landy-Szalay estimator (Landy & Szalay 1993), along with our fiducial band power estimates for the weak lensing signals.

Figure D.1 shows the resulting correlation matrix. Cross-correlations between the clustering correlation functions and any of the weak lensing signals only very rarely exceeds 0.05, and this is largely due to residual noise in the mock covariance. We can therefore safely assume that the clustering and weak lensing parts of the data vector are independent. This trend is driven by the fact that only 3 % of the BOSS survey area overlaps with KiDS and thus with the weak lensing measurements. Joint clustering and weak lensing measurements over the same sky area do produce significant cross-correlations (e.g. Krause & Eifler 2017; Krause et al. 2017) and therefore demand for a more homogeneous approach to summary statistics and their covariance than taken in this work.

Appendix D.2: Correlation function covariance

It is instructive to study the covariance models of the weak lensing correlation functions as an intermediate step to calculating
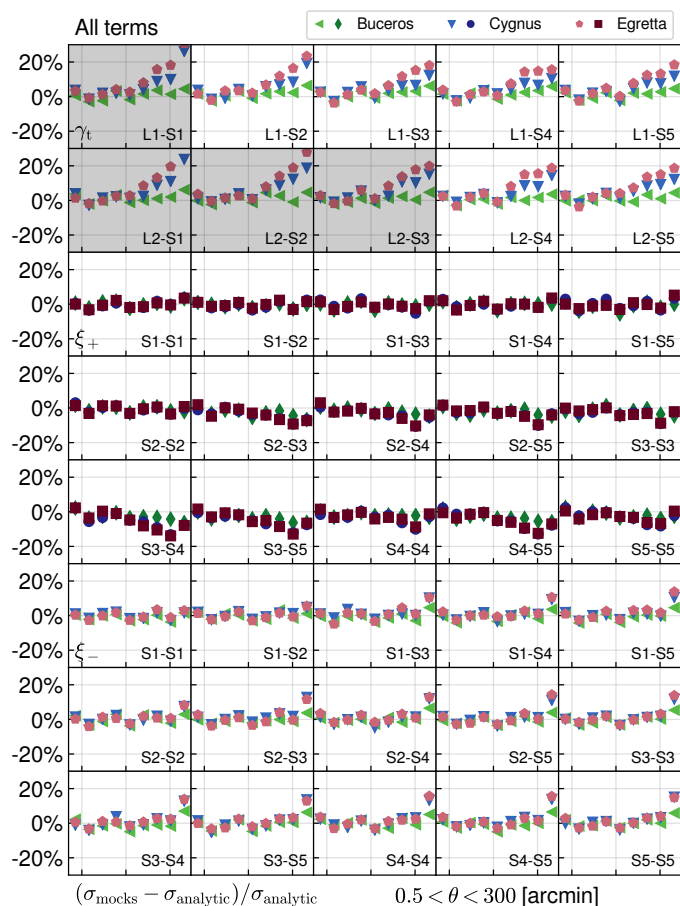
Fig. D.2. Relative difference between the square root of the diagonals of the mock and analytic covariances of the weak lensing correlation functions, i.e. the real-space analogue of Fig. 13. The top two rows show GGL signals, the centre three rows ξ_+, and the bottom three rows ξ_-, with bin combinations indicated in the panels. Three cases are shown: spatially uniform galaxy samples in a simple survey footprint (Buceros, green), spatially uniform galaxy samples in the realistic footprints (Cygnus, blue), and spatially varying samples in the realistic footprints (Egretta, red). GGL signals that are not used in the analysis have been greyed out.

band power covariance and as a view that isolates any deviations that are localised in configuration space. We compare analytic and mock covariances in the Buceros (simple rectangular survey geometry, uniform galaxy distributions), Cygnus (realistic survey footprints, uniform galaxy distributions), and Egretta (realistic footprint and spatially varying galaxy distributions) cases for tangential shear and the cosmic shear correlation functions $\xi_\pm$ measured in nine angular bins spread equidistantly in the log between 0.5 arcmin and 300 arcmin (i.e. the same scheme as used in KV450). In these comparisons the GGL estimator was applied to the mock catalogues with 100 times more random points than lens galaxies in order to suppress any residual contributions of terms that the random correction of $\langle \gamma_t \rangle$ removes (see Appendix E for details).

Figure D.2 shows the ratio of the square root of the diagonal elements of the mock and analytic covariances for the three survey configurations. Off-diagonals in the form of correlation coefficients are shown in Fig. D.3 for the most realistic Egretta case (the Fourier space analogues of these plots are Figs. 13 and 14). We generally find very good agreement between the mocks and the analytic approach in all cases. Significant devia-

tions are limited to the largest-scale data point in $\xi_\pm$ and to scales larger than 10 arcmin for GGL, with the mock standard deviations up to 20 % larger. This under-prediction by the analytic model is because its Gaussian sample variance terms ignore survey boundary effects which enhance them by up to a factor two when switching from a simple rectangular footprint to the realistic KiDS survey geometry (Buceros to Cygnus)[30]. Variable depth effects (Cygnus to Egretta) have negligible impact on the cosmic shear covariance beyond modifications to the galaxy pair counts, but cause a small additional increase in the GGL covariance.

To gain a better understanding of which covariance terms drive certain discrepancies, we create two special cases that are readily realised in both the analytic and simulation approaches. First, we remove all shear signals, which leaves us with only noise terms and, in the GGL case, with mixed terms that combine clustering with shape noise (referred to as the 'no-shear' case). Secondly, we set $\sigma_\epsilon = 0$ which removes all terms containing shape noise but keeps all Gaussian and non-Gaussian sample variance contributions, plus GGL mixed terms involving clustering shot noise (referred to as the 'signal-only' case). As discussed in Sect. 5.3, we do not expect the non-Gaussian covariance terms to match quantitatively between mock and analytic model. Hence, instead of directly comparing the two in the signal-only case, we investigate how well the two approaches agree in the changes between the different survey configurations.

The no-shear case in Fig. D.4 demonstrates excellent agreement in the cosmic shear noise terms; using the galaxy pair counts of the measurements in the analytic covariance correctly accounts for survey geometry and variable depth. The large-scale deviations are however still present in the GGL covariance and are caused by the residual mixed noise-sample variance contribution that in our model disregards survey boundary effects. These mixed terms, together with the Gaussian sample variance term which is of similar size (cf. Fig. 10), also drive the deviations seen in Fig. D.2. We note in passing that using Eq. (C.13) for the calculation of the GGL noise is critical for the high accuracy shown here. A naive simple pair count of lens and source galaxy pairs leads to 25 % difference in the standard deviations on small angular scales when the source and lens bins have substantial overlap.

For the interpretation of the signal-only case in Fig. D.5 it is helpful to bear in mind that the only contributing analytic covariance term that is sensitive to survey geometry is the super-sample covariance (SSC). As SSC is suppressed in GGL, the analytic model predicts no geometry dependency, which the mock result suggests is plausible above ca. 10 arcmin. Below however, non-Gaussian contributions are seen to cause excess covariance in the Egretta case, but it remains unclear to what degree this effect is influenced by the lognormal and linear galaxy bias assumptions, as well as the resolution limit, in our simulations. As regards the cosmic shear signals, mocks and analytic model generally agree fairly well in that the signal-only covariance is suppressed in the Egretta relative to the Buceros case[31], while on large scales dis-

---

[30] There is a further complication not included in the analytic model in that the GGL correlations include lenses beyond the limits of the source sample footprint. We employ the lens survey area in the analytic calculation of the sample variance terms, but the effective survey area is somewhat reduced because sources with gravitational shear estimates are not available over the full lens survey area. We estimate this to lead to a $2-3$ % under-prediction by the analytic model of the GGL standard deviation on large scales, which can therefore not be the main reason behind the discrepancies seen.

[31] While having the same total area, the Egretta footprint is spread out over a larger fraction of the sky, primarily due the separation between
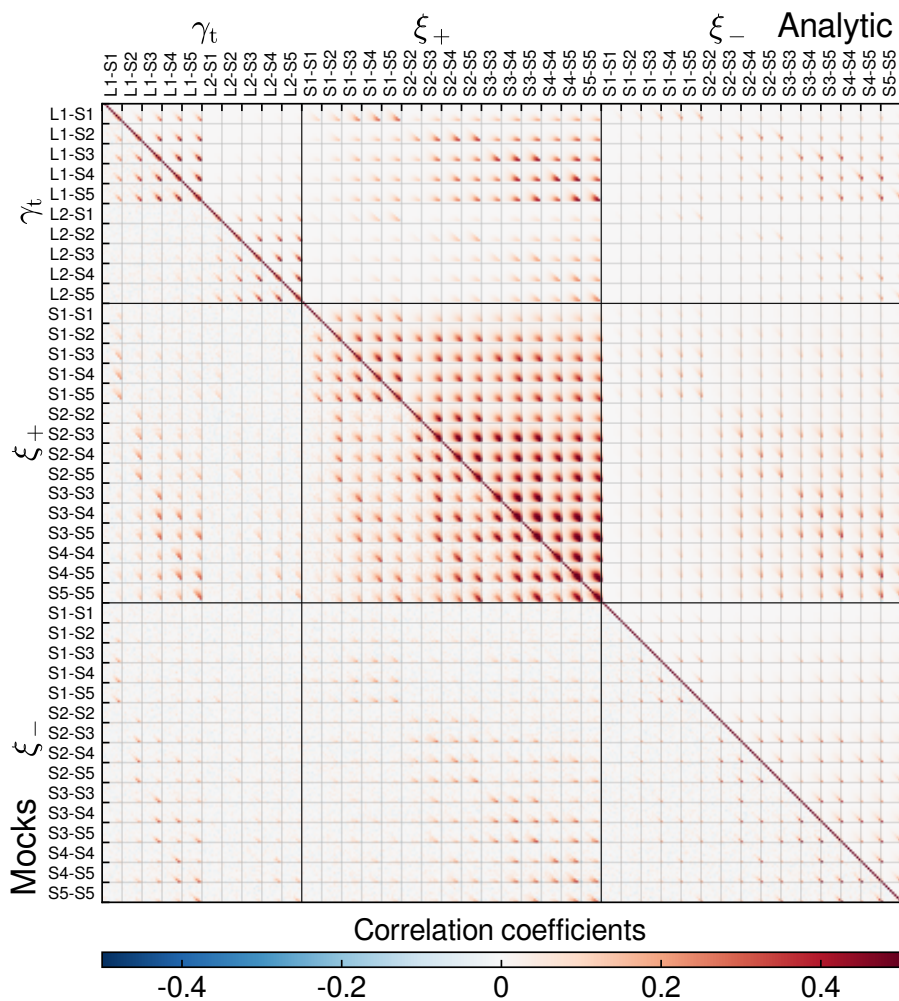
Fig. D.3. Comparison of the correlation coefficients in the weak lensing correlation function covariance between the mocks (lower left) and analytic approach (upper right) in the Egretta setup. As indicated in the labels, large blocks correspond to the correlation functions $\langle \gamma_t \rangle$, ξ_+, and ξ_-, while small blocks correspond to the tomographic bin combinations.

crepancies occur, again due to the neglect of survey geometry effects in the Gaussian sample variance. It is interesting to note that even in the signal-only case the differences in the standard deviation between analytic and mock covariances never exceed 30 %.

Patterns in the correlation structure away from the diagonal are well represented in the analytic covariance, with some small deviations discernible that, analogously to the Fourier space case, occur in terms dominated by Gaussian covariance contributions, but interestingly with a reversed sign, that is for correlation functions the analytic correlation coefficients are larger than their mock counterparts (compare Figs. D.3 and 14). For completeness, we also plot the relative difference between the Egretta and Buceros cases for the full covariance in Fig. D.6 (the analogue of Fig. 12). This figure confirms that the analytic covariance model overall performs very well in recovering the relevant effects of survey geometry and spatial variations. The small residual discrepancies, primarily on large angular scales, mirror those discussed in Fig. D.2.

Appendix D.3: Band power covariance

Here we provide additional band power covariance comparison plots that further illustrate the fidelity and importance of individual contributions to the covariance model: Fig. D.7 compares the mock and analytic approaches in the no-shear case; Fig. D.8 shows the relative difference between the Egretta and Buceros setups in the signal-only case. The observed trends are in line with the configuration space covariances. The noise terms agree well, whereas the idealised mixed sample variance-noise terms in the analytic model fail to capture survey geometry effects on large scales in the GGL covariance. While the analytic model suggests minimal impact on sample variance in GGL due to survey geometry and/or variable survey depth, the mocks yield a $10 - 20\%$ excess standard deviation switching from Buceros to Egretta, but it is not clear if this is physical or due to limitations in the mock creation. The trend is reversed for cosmic shear, driven by SSC. Mocks and analytic prediction agree on the scale dependence of this effect, with the latter slightly over-predicting its magnitude.

Figure D.9 highlights the differences in correlation coefficients between the mock and analytic covariance in the Egretta case. For the vast majority of coefficients the differences are consistent with scatter due to the finite number of mock realisations. Systematic differences lie in the very thin tails, mostly at the positive end, indicating a larger mock value. These typically occur in the cross-variance between cosmic shear and GGL and are dominated by the mixed term.

KiDS-N and KiDS-S. Therefore, the Egretta footprint is able to accommodate larger modes of the large-scale matter density distribution, thereby decreasing super-sample covariance.
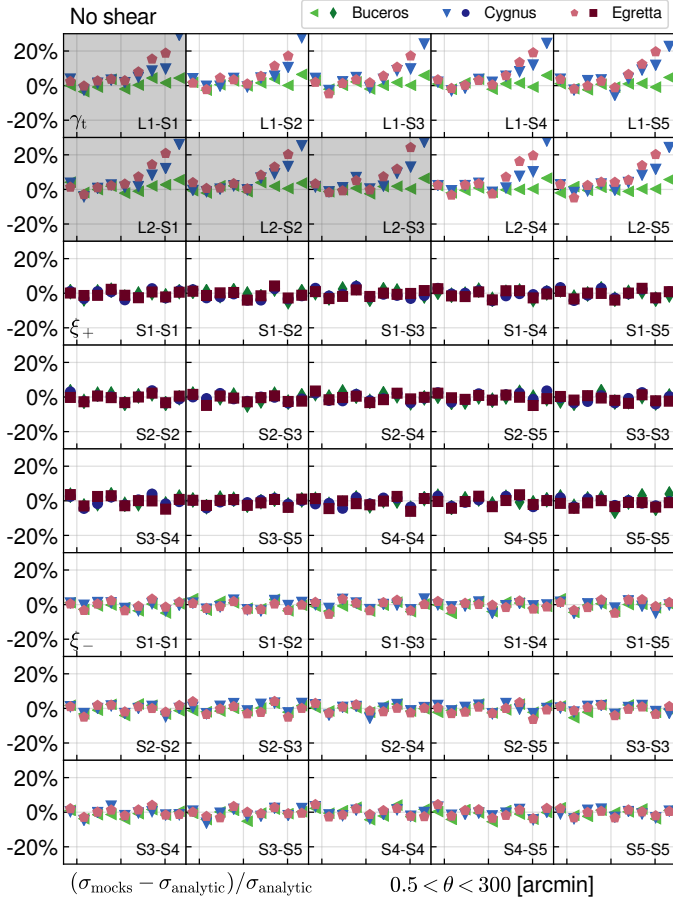
Fig. D.4. Same as Fig. D.2, but with all weak lensing signals removed from both the analytic and mock covariances, i.e. only noise and sample variance due to clustering contribute.
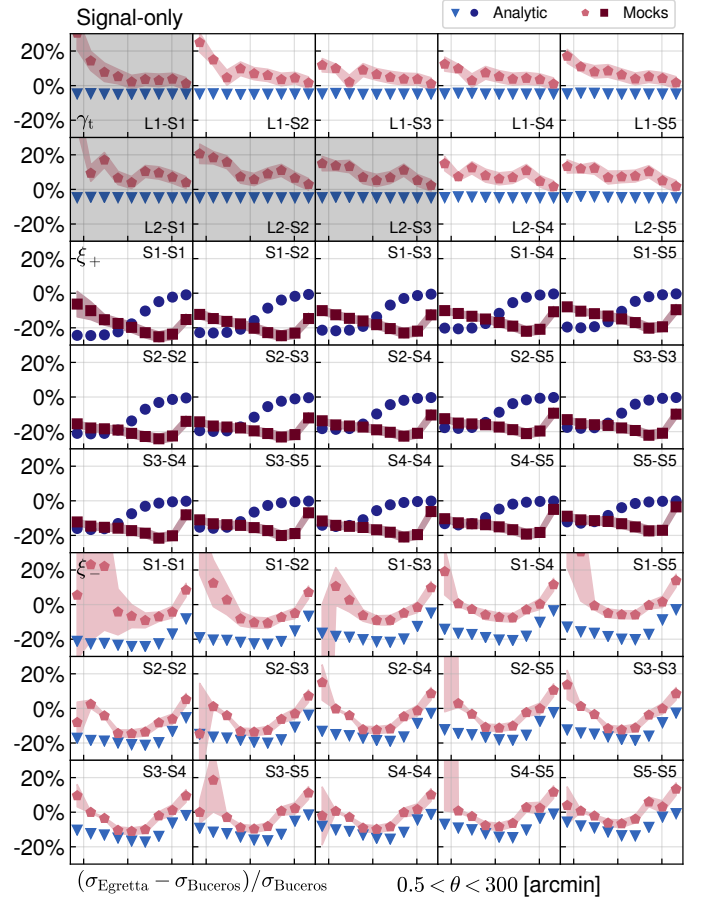
Fig. D.5. Relative difference between the square root of the diagonals in the Egretta (realistic mask and depth variations) and Buceros (rectangular mask, uniform depth) covariances of the weak lensing correlation functions, with all shape noise contributions removed ($\sigma_\epsilon = 0$). Red (blue) symbols show results for the mock (analytic) covariance. Bands around the mock data points indicate the standard error determined from a jackknife estimate of variance. The top two rows show GGL signals, the centre three rows ξ_+, and the bottom three rows ξ_-, with bin combinations indicated in the panels.
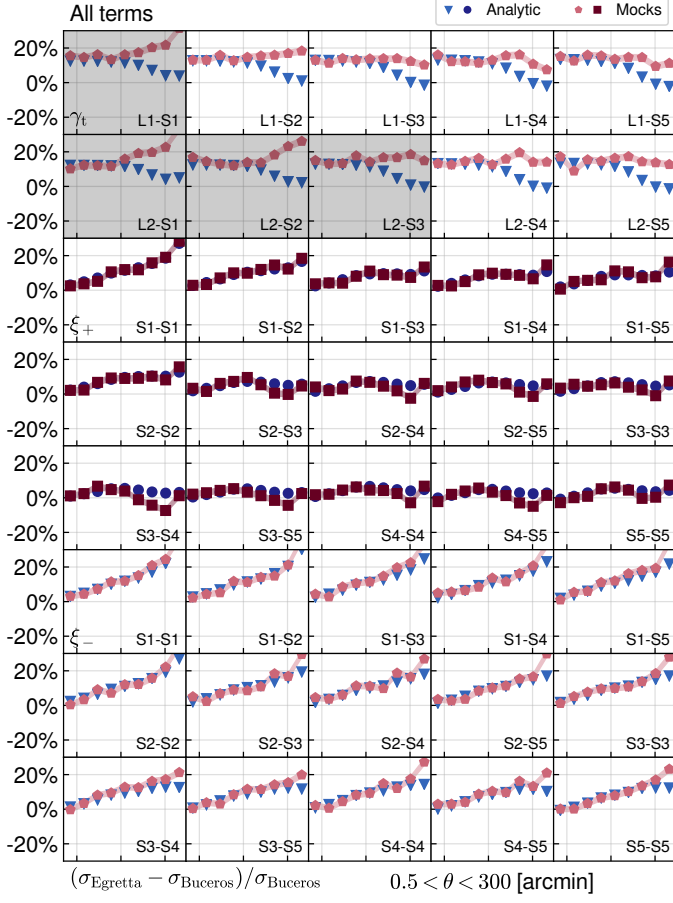
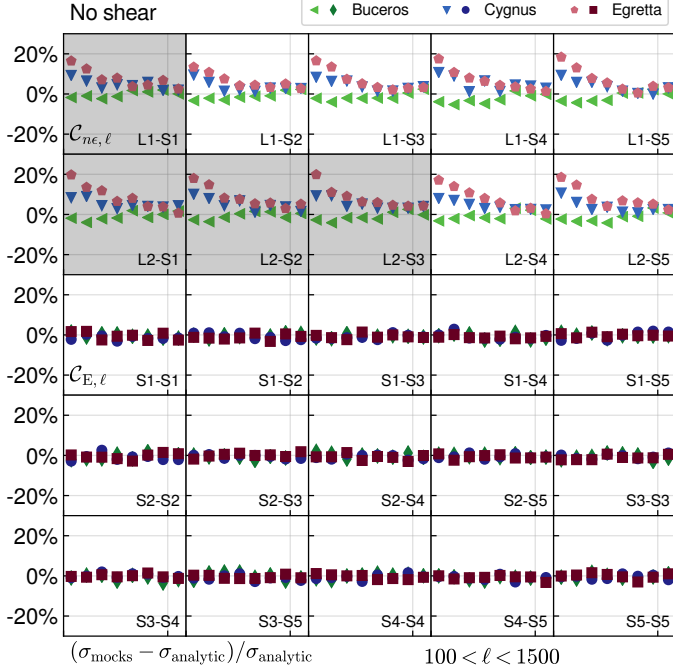**Fig. D.6.** Same as Fig. D.5, but for the full covariance with all cosmological contributions included.

Fig. D.7. Same as Fig. 13, but with all weak lensing signals removed from both the analytic and mock covariances, i.e. only noise and sample variance due to clustering contribute.
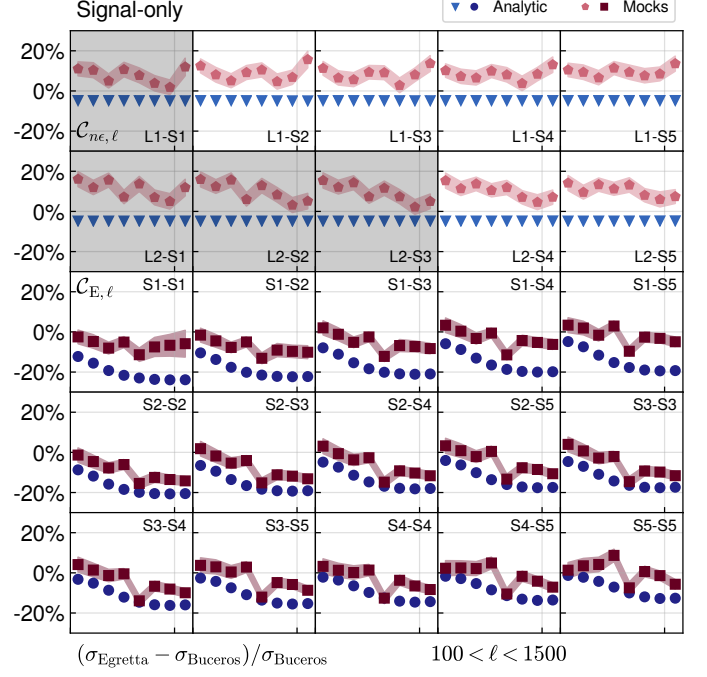
Fig. D.8. Same as Fig. 12, but with all shape noise contributions to the analytic and mock covariances removed ($\sigma_\epsilon = 0$).

Fig. D.9. Relative difference between the correlation coefficients of the mock and analytic covariances of the weak lensing band power signals. *Bottom*: Histogram of the relative differences marking values beyond $\pm 6\,\%\,(\pm 3\,\%)$ in dark (light) red and blue. *Top*: Band power correlation matrix with elements coloured according to the value of the difference using the same scheme as in the matrix above.

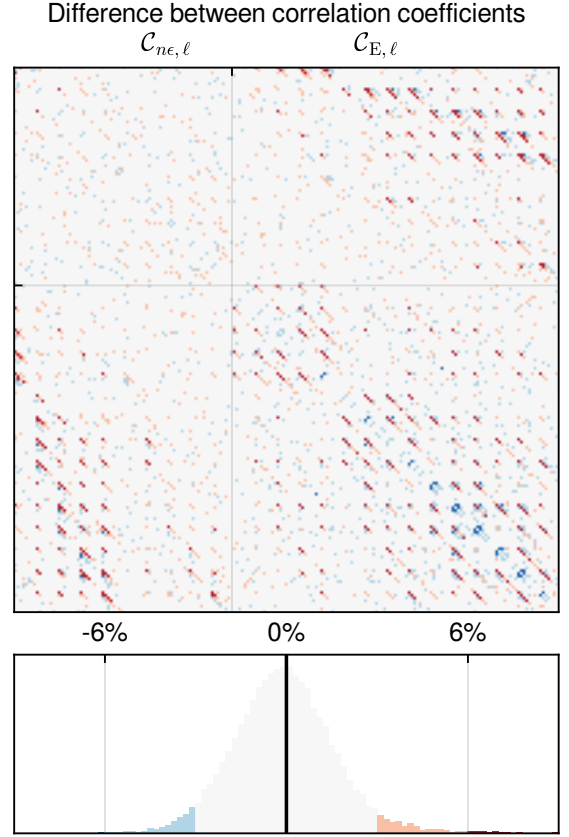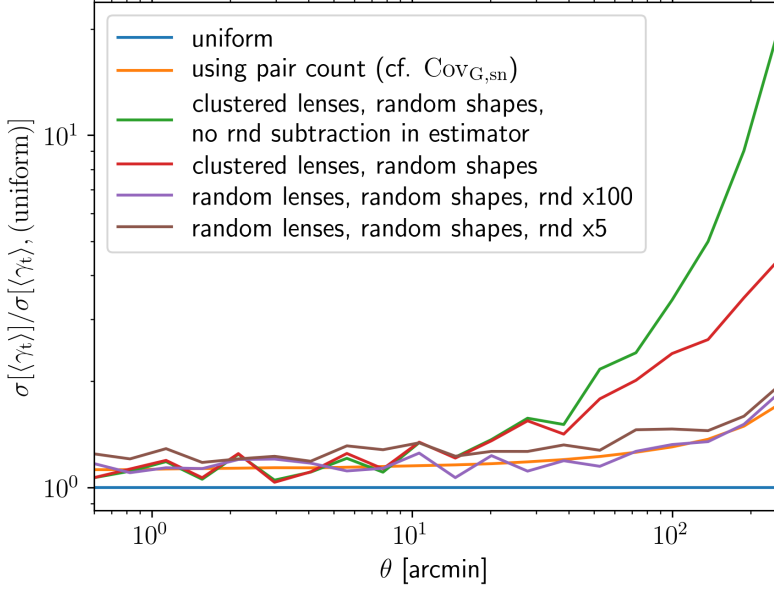**Fig. E.1.** Noise contributions to a GGL tangential shear correlation function measurement for BOSS lenses in the KiDS-1000 overlap (L1) and KiDS high-redshift sources (S5). Shown is the standard deviation as a function of angular separation, normalised by the analytic expectation for uniformly distributed galaxies and neglecting survey boundaries (blue). The noise term as implemented by Eq. (E.3) is shown in orange. It is in excellent agreement with the sample variance obtained from a measurement with randomised source galaxy shapes and lens galaxy positions, i.e. removing all sample variance contributions (purple). Decreasing the oversampling factor of the random catalogue from our default of $\mathcal{N}_{\mathrm{rnd}} = 100$ to only five results in an increased noise level (brown). We also show the signals when only randomising shapes (red) and when additionally not subtracting the GGL signal around random points in the lens sample (green).

Appendix E: Analytic covariance model

Here we provide a detailed description of our analytic covariance model. This model is used for the cosmic shear and galaxy-galaxy lensing (GGL) signals only, while the clustering covariance is obtained from the public mocks provided by BOSS. With our own mocks we demonstrate that the cross-variance between the clustering and lensing observables can safely be neglected. For completeness we include the analytic expressions for angular clustering as well. We calculate real-space correlation function covariances first, and then obtain the band power covariance from these expressions analogously to how the signals are derived. While in principle it is possible to go directly from angular power spectrum models to band powers following Eqs. (25,32), we choose this approach for two reasons: first, we already have validated correlation function covariance software in hand from previous KiDS cosmology analyses (Hildebrandt et al. 2017; van Uitert et al. 2018; Hildebrandt et al. 2020a), and secondly, it is easier to incorporate survey effects in real space, such as measured galaxy pair counts.

Appendix E.1: Gaussian real-space covariance

We begin with the Gaussian covariance, which corresponds to the full contribution if the underlying gravitational lensing convergence and galaxy number density fields were Gaussian. It consists of a sample variance term ('sva'), the sampling error due to observing a finite volume of the Universe, shape noise or shot noise ('sn'), as we observe galaxies as point processes sampling the underlying fields, and a mixed term ('mix'). We opt to derive the real-space sample variance expression from its Fourier counterpart as proposed by Joachimi et al. (2008), leading to

$$\mathrm{Cov}_{\mathrm{G,sva}}\left[\Xi_{\mu}^{(ij)}(\bar{\theta}_1); \Xi_{\nu}^{(kl)}(\bar{\theta}_2)\right] = \frac{1}{2\pi A_{\mathrm{max},\mu\nu}} \int_0^\infty \mathrm{d}\ell\,\ell\,\mathcal{K}_\mu(\ell\bar{\theta}_1)\mathcal{K}_\nu(\ell\bar{\theta}_2)\left\{C^{(ik)}(\ell)C^{(jl)}(\ell) + C^{(il)}(\ell)C^{(jk)}(\ell)\right\}\,, \tag{E.1}$$

where we introduced a unified notation for correlation functions with the correspondence $\{w, \langle\gamma_t\rangle, \xi_+, \xi_-\} \leftrightarrow \{\Xi_0, \Xi_2, \Xi_0, \Xi_4\}$. The subscripts denote the type of integration kernel that is applicable, with ξ_+ and the angular clustering correlation function w sharing the same kernel. The kernels are defined as

$$\mathcal{K}_\mu(\ell\bar{\theta}_i) := \frac{2}{\theta_{\mathrm{u},i}^2 - \theta_{\mathrm{l},i}^2}\int_{\theta_{\mathrm{l},i}}^{\theta_{\mathrm{u},i}}\mathrm{d}\theta'\,\theta'\mathrm{J}_\mu(\ell\theta') = \frac{2}{(\theta_{\mathrm{u},i}^2 - \theta_{\mathrm{l},i}^2)\ell^2} \times \begin{cases} [x\mathrm{J}_1(x)]_{\ell\theta_{\mathrm{l},i}}^{\ell\theta_{\mathrm{u},i}} & \mu = 0 \\ [-x\mathrm{J}_1(x) - 2\mathrm{J}_0(x)]_{\ell\theta_{\mathrm{l},i}}^{\ell\theta_{\mathrm{u},i}} & \mu = 2 \\ \left[\left(x - \frac{8}{x}\right)\mathrm{J}_1(x) - 8\mathrm{J}_2(x)\right]_{\ell\theta_{\mathrm{l},i}}^{\ell\theta_{\mathrm{u},i}} & \mu = 4 \end{cases}\,, \tag{E.2}$$

where the J_μ are cylindrical Bessel functions of the first kind. As opposed to earlier works, we have explicitly averaged over the angular bin centred on $\bar{\theta}_i$ in which Ξ_μ is measured, delimited by $[\theta_{\mathrm{l},i}; \theta_{\mathrm{u},i}]$. The type of angular power spectrum $C^{(ij)}(\ell)$ to be used in the integrand of Eq. (E.1) is determined by its tomographic bins i and j: if both are lens bins, it is a clustering power spectrum; if both are source bins, it is a cosmic shear power spectrum; else one uses a position-shear cross power spectrum. For a more explicit notation in this regard see Joachimi & Bridle (2010). The power spectra are determined under the extended Limber approximation from the full non-linear matter power spectrum evaluated at our fiducial choice of parameters. The power spectra are calculated using the non-linear prescription by Takahashi et al. (2012) as AGN feedback effects have negligible impact on the covariance in the regime where sample variance contributions are significant (see also Schneider et al. 2020). Lensing signals generally include intrinsic alignment contributions in the Gaussian terms, but they have been switched off for this study to simplify the comparison with the mocks that do not feature intrinsic alignments. We assume an effective linear galaxy bias and choose the fiducial values of b_1 for this purpose as the corrections due to the non-linear bias terms are small.

Following the derivation in van Uitert et al. (2018), we normalise the covariance term by $A_{\mathrm{max},\mu\nu} = \max(A_{\mathrm{eff},\mu}, A_{\mathrm{eff},\nu})$, that is the effective survey area applicable to the signal in the case of auto-correlations and the *larger* effective area when cross-correlating signals measured over different parts of the sky. The areas to be used are the full BOSS footprint for clustering, the full KiDS-1000 area for cosmic shear, and the overlap area for GGL; see Table 1 for numerical values. The effective survey area is not a quantity defined from first principles and ultimately depends on the chosen resolution at which the survey footprint is considered. We measure the effective area from a binary HEALPIX mask with $N_{\mathrm{side}} = 4096$. Hence, mask features of less than arcminute size will not reduce A_{eff}. Since we use this area in n_{eff} as well (see Eq. C.12), star masks and other small-scale features are interpreted as diluting the number density of galaxies rather than the survey area. This is in line with the covariance modelling assumptions as long as these small-scale features are below the scales at which the cosmological signal is measured. We have also computed a cosmic shear covariance based on an $N_{\mathrm{side}} = 2048$ mask, but found no measurable difference on S_8 best fits or its errors.

The pure noise term only contributes to the diagonals of auto-correlations and is given by (Schneider et al. 2002a)

$$\mathrm{Cov}_{\mathrm{G,sn}}\left[\Xi_\mu^{(ij)}(\bar\theta_1); \Xi_\nu^{(kl)}(\bar\theta_2)\right] = \delta_{\bar\theta_1\bar\theta_2}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)\frac{\mathcal{T}_{\mu\nu}^{\mathrm{sn}}}{N_{\mathrm{pair}}^{(ij)}(\bar\theta_1)}\ , \tag{E.3}$$

where δ_{ij} denotes a Kronecker delta, and where $N_{\mathrm{pair}}^{(ij)}$ is given by Eqs. (C.11) and (C.13). We defined

$$\mathcal{T}_{\mu\nu}^{\mathrm{sn}} := \begin{cases} \sigma_\epsilon^4/2 & \mu = \nu = 0 \text{ or } \mu = \nu = 4 \quad \text{(cosmic shear)} \\ \sigma_\epsilon^2/2 & \mu = \nu = 2 \qquad\qquad\qquad \text{(GGL)} \\ 1 & \mu = \nu = 0 \qquad\qquad\qquad \text{(clustering)} \\ 0 & \mu \neq \nu \end{cases}, \tag{E.4}$$

where σ_ϵ is the total dispersion of the complex observed galaxy ellipticity (with contributions from the intrinsic ellipticities of galaxies and measurement noise) that is in practice measured via Eq. (C.9). It was furthermore assumed that the noise in galaxy clustering follows a Poisson distribution. The mixed term reads

$$\mathrm{Cov}_{\mathrm{G,mix}}\left[\Xi_\mu^{(ij)}(\bar\theta_1); \Xi_\nu^{(kl)}(\bar\theta_2)\right] = \delta_{jl}\frac{\mathcal{T}_j^{\mathrm{mix}}}{2\pi n_{\mathrm{eff}}^{(j)}A_{\mathrm{max},\mu\nu}}\int_0^\infty \mathrm{d}\ell\,\ell\,\mathcal{K}_\mu(\ell\bar\theta_1)\mathcal{K}_\nu(\ell\bar\theta_2)\,C^{(ik)}(\ell) + 4\text{ perm.}\ , \tag{E.5}$$

where we defined

$$\mathcal{T}_j^{\mathrm{mix}} := \begin{cases} \sigma_\epsilon^2/2 & j \in \mathrm{S1}-\mathrm{S5} \\ 1 & j \in \mathrm{L1}-\mathrm{L2} \end{cases} \tag{E.6}$$

for our lens (L) and source (S) bins, respectively. Figure E.1 illustrates the noise contributions to the variance of an exemplary GGL tangential shear signal (L1–S5). Sample variance contributions are switched off selectively by randomising the orientations of source galaxy shapes and/or the positions of lens galaxy positions.

First, the measured sample variance is offset by 13 % over the expectation for uniformly distributed galaxies in the absence of any survey boundaries for which $N_{\mathrm{pair}}^{(ij)}(\theta_i) = \pi(\theta_{\mathrm{u},i}^2 - \theta_{\mathrm{l},i}^2)A_{\mathrm{eff}}n_{\mathrm{eff}}^{(i)}n_{\mathrm{eff}}^{(j)}$ is inserted into Eq. (E.3). This is an example of the aforementioned importance of the choice of effective survey area. We used the BOSS area in the calculation which is larger than the source area due to the more aggressive masking in weak lensing surveys, thereby underestimating the true noise even on small scales where survey boundaries are irrelevant. Our noise term model closely follows the measured GGL standard deviation without any sample variance (i.e. randomised lenses and sources). An oversampling factor of 100 of the random catalogue is sufficient to suppress any additional noise contributions, which we therefore use as our default.

Only randomising the sources preserves a mixed term in the resulting variance, namely a combination of shape noise and angular clustering signal; cf. Eq. (E.5). Figure E.1 shows that this term dominates noise components on large scales, so that idealisations in the modelling of this term will limit the accuracy of the analytic model, as can be seen in Fig. D.4. Moreover, the figure demonstrates that the subtraction of the GGL signal around random lens points in the estimator (Eq. 38) is crucial to suppress additive noise contributions that scale with shape noise and the survey footprint of the lens sample, which would otherwise substantially increase errors on large scales (see the detailed discussion in Singh et al. 2017). Equation (E.3) was previously shown to be an excellent fit to the noise term of cosmic shear signals; see Fig. D.4 and Troxel et al. (2018a).

The ostensibly simplest component, the Gaussian sample variance contribution, currently limits the accuracy of this model (see Appendix D). It dominates on the largest scales where the finite extent of the survey footprint affects the sample variance substantially and breaks the assumption of isotropy; see also the discussion in Blake et al. (2020). Eqs. (E.1) and (E.5) acquire their simple form by neglecting survey boundaries altogether. In future, these effects, as well as variations in survey depth, will be easier to account for in a configuration space approach; see Kilbinger & Schneider (2004); Hikage et al. (2019).

Appendix E.2: Non-Gaussian real-space contributions

Since the galaxy density and weak lensing convergence distributions are highly non-Gaussian on small scales, the covariance picks up additional terms via the connected four-point function of these fields. These are conveniently split into matter trispectrum contributions from modes within the survey footprint and those that link in-survey modes to those with wavelengths larger than the survey which act to rescale the mean of the field inside the survey. The former are given by (Takada & Jain 2004)

$$\mathrm{Cov}_{\mathrm{NG}}\left[\Xi_\mu^{(ij)}(\bar\theta_1); \Xi_\nu^{(kl)}(\bar\theta_2)\right] = \frac{1}{4\pi^2 A_{\mathrm{max},\mu\nu}}\int_0^\infty \mathrm{d}\ell_1\,\ell_1\,\mathcal{K}_\mu(\ell_1\bar\theta_1)\int_0^\infty \mathrm{d}\ell_2\,\ell_2\,\mathcal{K}_\nu(\ell_2\bar\theta_2)\int_0^\pi \frac{\mathrm{d}\varphi_\ell}{\pi}\,T^{(ijkl)}(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,-\boldsymbol{\ell}_1,-\boldsymbol{\ell}_2)\ , \tag{E.7}$$

where φ_ℓ is the angle between ℓ_1 and ℓ_2. The angular trispectrum is calculated by integrating along the line of sight over the matter trispectrum, $T_{\rm m}$, yielding

$$T^{(ijkl)}(\ell_1, \ell_2, \ell_3, \ell_4) = \int_0^{\chi_{\rm hor}} {\rm d}\chi \; \frac{W_{\rm a}^{(i)}(\chi)\, W_{\rm b}^{(j)}(\chi)\, W_{\rm c}^{(k)}(\chi)\, W_{\rm d}^{(l)}(\chi)}{f_{\rm K}^6(\chi)} \; T_{\rm m}\left(\frac{\ell_1}{f_{\rm K}(\chi)}, \frac{\ell_2}{f_{\rm K}(\chi)}, \frac{\ell_3}{f_{\rm K}(\chi)}, \frac{\ell_4}{f_{\rm K}(\chi)}\right) \tag{E.8}$$

under Limber's approximation. The type of kernel $W_{a,b,c,d}$ is chosen according to the probes Ξ under consideration from among Eq. (15) for weak lensing and the comoving distance distribution of lens samples plus an expression for galaxy bias for clustering.

The super-sample covariance (SSC) term reads (Takada & Hu 2013)

$$\mathrm{Cov}_{\rm SSC}\left[\Xi_\mu^{(ij)}(\bar{\theta}_1); \Xi_\nu^{(kl)}(\bar{\theta}_2)\right] = \frac{1}{4\pi^2} \int_0^\infty {\rm d}\ell_1\, \ell_1\, \mathcal{K}_\mu(\ell_1 \bar{\theta}_1) \int_0^\infty {\rm d}\ell_2\, \ell_2\, \mathcal{K}_\nu(\ell_2 \bar{\theta}_2) \tag{E.9}$$

$$\times \int_0^{\chi_{\rm hor}} {\rm d}\chi\; \frac{W_{\rm a}^{(i)}(\chi)\, W_{\rm b}^{(j)}(\chi)\, W_{\rm c}^{(k)}(\chi)\, W_{\rm d}^{(l)}(\chi)}{f_{\rm K}^6(\chi)}\; \frac{\partial P_{\rm m}\left[\ell_1/f_{\rm K}(\chi)\right]}{\partial \delta_{\rm b}}\; \frac{\partial P_{\rm m}\left[\ell_2/f_{\rm K}(\chi)\right]}{\partial \delta_{\rm b}}\; \sigma_{{\rm bg},\mu\nu}^2(\chi)\,,$$

where the derivatives denote the response of the matter power spectrum to a change in the density contrast of the background $\delta_{\rm b}$, which is defined as the average density contrast within the volume of the survey. Here, we defined the variance of background matter fluctuations within the observability masks relevant to the two probes under consideration (indicated by super-/subscripts μ and ν),

$$\sigma_{{\rm bg},\mu\nu}^2(\chi) = \frac{1}{A_{{\rm eff},\mu}\, A_{{\rm eff},\nu}} \sum_\ell P_{\rm m,lin}\left(\frac{\ell}{f_{\rm K}(\chi)}\right) \sum_m a_{\ell m}^\mu\, a_{\ell m}^{\nu *}\,, \tag{E.10}$$

where the linear matter power spectrum has been used as only linear scales affect the background fluctuations. As surveys now cover substantial fractions of the sky, we drop the flat-sky approximation in this term and express the cross-power of the survey masks via their spherical harmonic coefficients $a_{\ell m}^{\mu,\nu}$. These are determined from the same binary HEALPIX masks used to calculate the effective survey area. For probes covering the same sky area the summation over m simplifies to the isotropic power spectrum of the mask. We do not include contributions caused by super-survey tidal fields, which are expected to have little impact on our scales of interest but could attain similar levels as the (small) NG contribution (Barreira et al. 2018b).

To evaluate the matter trispectrum and the matter power spectrum response to the background we opt for a halo model formalism, closely following Takada & Hu (2013); Li et al. (2014); see also Krause & Eifler (2017) for a similar implementation. We refer the reader to these works for a quantitative description and only summarise the relevant modelling choices here, which are unchanged with respect to earlier KiDS cosmology analyses (Hildebrandt et al. 2017; van Uitert et al. 2018; Hildebrandt et al. 2020a). Our halo model is based on the halo mass function and halo bias of Tinker et al. (2010). It assumes a Navarro et al. (1996) halo profile with the concentration-mass relation by Duffy et al. (2008) and employs the analytical form of the profile's Fourier transform by Scoccimarro et al. (2001). While implemented, some of the particularly computationally expensive 2-halo terms in the matter trispectrum have been switched off for most practical covariance calculations because they only make negligible contributions to the non-Gaussian covariance, which has little impact on the overall statistical errors to begin with. The logarithmic matter power spectrum response is reduced by two for both clustering (de Putter et al. 2012; Takada & Hu 2013) and GGL (Singh et al. 2017) signals as their estimators are normalised to the mean galaxy densities within the survey footprint rather than the global mean density through the use of random catalogues. An effective linear bias is used to translate from polyspectra of matter to those involving galaxy density contrast, and we choose the fiducial values of b_1 for this purpose.

Appendix E.3: Covariance of band powers

Together with the covariance term sourced through uncertainty in the multiplicative bias correction (Eq. 37), the expressions above constitute the full covariance for angular large-scale structure correlation functions. Various useful two-point statistics can be derived from the correlation functions, which capitalises on the insensitivity of the latter to the often complex survey geometry. Since the relations are linear, simple error propagation allows us to derive corresponding relations between the covariances of these two-point statistics. For the case of band powers, employing Eqs. (39) and (40) leads to

$$\mathrm{Cov}\left[C_{{\rm E/B},m}^{(ij)}; C_{{\rm E/B},n}^{(kl)}\right] = \frac{\pi^2}{N_m\, N_n} \sum_{a,b} \Delta\theta_a\, \theta_a\, T(\theta_a)\, \Delta\theta_b\, \theta_b\, T(\theta_b) \left\{ g_+^m(\theta_a)\, g_+^n(\theta_b)\, \mathrm{Cov}\left[\xi_+^{(ij)}(\theta_a); \xi_+^{(kl)}(\theta_b)\right] \right. \tag{E.11}$$

$$\left. + g_-^m(\theta_a)\, g_-^n(\theta_b)\, \mathrm{Cov}\left[\xi_-^{(ij)}(\theta_a); \xi_-^{(kl)}(\theta_b)\right] \pm g_+^m(\theta_a)\, g_-^n(\theta_b)\, \mathrm{Cov}\left[\xi_+^{(ij)}(\theta_a); \xi_-^{(kl)}(\theta_b)\right] \pm g_-^m(\theta_a)\, g_+^n(\theta_b)\, \mathrm{Cov}\left[\xi_-^{(ij)}(\theta_a); \xi_+^{(kl)}(\theta_b)\right] \right\}$$

for cosmic shear, and

$$\mathrm{Cov}\left[C_{{\rm n}\epsilon,m}^{(ij)}; C_{{\rm n}\epsilon,n}^{(kl)}\right] = \frac{4\pi^2}{N_m\, N_n} \sum_{a,b} \Delta\theta_a\, \theta_a\, T(\theta_a)\, \Delta\theta_b\, \theta_b\, T(\theta_b)\, h^m(\theta_a)\, h^n(\theta_b)\, \mathrm{Cov}\left[\langle\gamma_{\rm t}\rangle^{(ij)}(\theta_a); \langle\gamma_{\rm t}\rangle^{(kl)}(\theta_b)\right]\,, \tag{E.12}$$

for GGL. The cross-variances between these signals, and expressions involving clustering (not used in this work), are obtained analogously. The angular binning is the same as that of the correlation functions, which is described in Sect. 3.6.