A&A 646, A129 (2021)
https://doi.org/10.1051/0004-6361/202038831

Astronomy & Astrophysics

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, PO 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, PO Box 9513, 2300 RA 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, Zhuhai Campus, Guangzhou 519082, PR China
[14] Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA
[15] Shanghai Astronomical Observatory (SHAO), Nandan Road 80, Shanghai 200030, PR China
[16] University of Chinese Academy of Sciences, Beijing 100049, PR China

Received 3 July 2020 / Accepted 17 December 2020

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 $S_8 = \sigma_8\,(\Omega_\mathrm{m}/0.3)^{1/2}$ 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. cosmology: miscellaneous – gravitational lensing: weak – large-scale structure of Universe – methods: data analysis – methods: analytical – methods: statistical

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 $8\,h\,\mathrm{Mpc}^{-1}$ (where $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$), 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 VI 2020; Planck Collaboration V 2020).

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. 2020, 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. 2018a). 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. 2020a). 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. 2020; 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. 2020a).

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 $1006\,\mathrm{deg}^2$ out of the final KiDS area of $1350\,\mathrm{deg}^2$. 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. (2020a) 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: Sect. 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. 2021) and redshift distributions (Hildebrandt et al. 2021), the cosmological analysis of cosmic shear (Asgari et al. 2021a) and of joint clustering and weak lensing (Heymans et al. 2021), and an investigation of extensions to the spatially flat ΛCDM model considered throughout here (Tröster et al. 2020b). 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 $k \sim 0.3\,h\,\mathrm{Mpc}^{-1}$, 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 $k > 1\,h\,\mathrm{Mpc}^{-1}$, 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 hydrodynamical 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. 2021b; Park et al. 2021 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. 2020), 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. (2021a).

2.1. Matter power spectra

Linear matter power spectra $P_{\mathrm{m,lin}}(k, z)$ 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, $\sum m_\nu = 0.06$ eV. We keep this parameter fixed in our default setup, in line with Planck Collaboration VI (2020), and consider constraints on neutrino mass in a companion paper (Tröster et al. 2020b).

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 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. 2020); this remains the case for KiDS-1000 (Asgari et al. 2021a). 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 Ω_{m}–σ_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

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

[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.

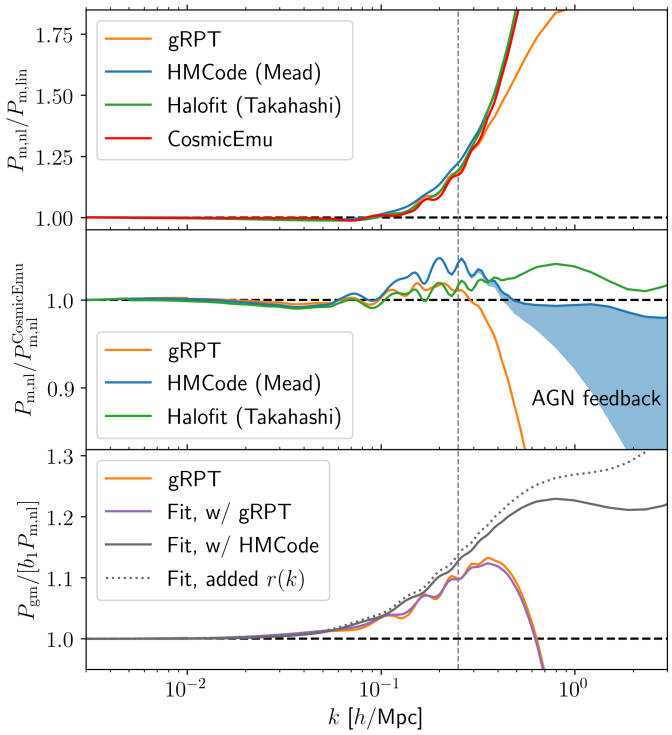

Fig. 1. Comparison of 3D power spectra, computed at $z = 0.38$ and the parameters listed in Table A.1. *Top*: non-linear matter power spectra relative to the linear matter power spectrum, for the gRPT perturbative model, the Mead et al. (2015) HMCODE, the Takahashi et al. (2012) Halofit prescription, 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 $P_{\rm gm}$ 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.

Takahashi et al. (2012) Halofit model as we limit the feedback model to only account for suppression of power.

Recently, Euclid Collaboration (2019) introduced the EuclidEmulator and demonstrated it to be in excellent agreement with CosmicEmu to ∼1% at $z \lesssim 1$ and over the majority of relevant k-scales, despite using completely independent simulations and emulation methods[3]. The emulators agree that the Takahashi et al. (2012) Halofit model consistently overpredicts 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 2019). We can therefore reject Halofit as an accurate description of matter power at highly non-linear 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.

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. 2018a) and HSC (Hikage et al. 2019) cosmic shear analyses. However, Troxel et al. (2018a) 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. (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

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

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:

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

with the definition

$$\mathcal{F}_\alpha(k) := \frac{P_\alpha(k,z_{\rm ref})}{P^2_{\rm m,lin}(k,z_{\rm ref})} = \frac{h^3}{{\rm Mpc}^3} \exp\left\{ \sum_{i=0}^{2} f_{\alpha,i} \left[\ln\left(k \frac{\rm Mpc}{h} \right) \right]^i \right\} \tag{4}$$

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 $P^2_{\rm m,lin}$. 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

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

The complete set of best-fit coefficients $g^{mn}_{\alpha,i}$ 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. 2021b). 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. (2021b), 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. (2021b) 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 $k < 0.3\,h\,{\rm Mpc}^{-1}$, 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. 2021), 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

$$P_{\rm gg,s}(k,\mu,z) = W_\infty(f_{\rm g}\,k\,\mu) \Big\{ P_{\rm gg}(k,z) + 2 f_{\rm g}(a)\,\mu^2 P_{{\rm g}\theta}(k,z) + f_{\rm g}(a)^2 \mu^4 P_{\theta\theta}(k,z) + \text{higher-order terms} \Big\}, \tag{6}$$

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

$$f_{\rm g}(a) = \frac{{\rm d}\ln D(a)}{{\rm d}\ln a} \tag{7}$$

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

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

where $P_{{\rm m}\theta}(k,z)$ 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

[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.

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

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

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 $a_{\rm vir}$ 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,

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

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

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

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

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

for the perpendicular and line-of-sight components of the separation vector, respectively. We use the comoving angular diameter distance $f_{\rm K}$, the comoving (radial) distance χ, and the Hubble parameter $H(z)$ evaluated at the mean redshifts $z_{\rm mean}$ 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 $\Omega_{\rm m}$ (S17; Tröster et al. 2020a).

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

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 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

$$C^{(ij)}_{\epsilon\epsilon}(\ell) = C^{(ij)}_{\rm GG}(\ell) + C^{(ij)}_{\rm GI}(\ell) + C^{(ij)}_{\rm IG}(\ell) + C^{(ij)}_{\rm II}(\ell), \tag{13}$$

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

$$C^{(ij)}_{\rm ab}(\ell) = \int_0^{\chi_{\rm hor}} {\rm d}\chi\,\frac{W_a^{(i)}(\chi)\,W_b^{(j)}(\chi)}{f_{\rm K}^2(\chi)}\,P_{\rm m,nl}\left(\frac{\ell+1/2}{f_{\rm K}(\chi)},z(\chi)\right), \tag{14}$$

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

$$W^{(i)}_{\rm G}(\chi) = \frac{3H_0^2\Omega_{\rm m}}{2\,c^2}\,\frac{f_{\rm K}(\chi)}{a(\chi)}\int_\chi^{\chi_{\rm hor}} {\rm d}\chi'\,n^{(i)}_{\rm S}(\chi')\,\frac{f_{\rm K}(\chi'-\chi)}{f_{\rm K}(\chi')}, \tag{15}$$

where $\Omega_{\rm m}$ is the total matter density parameter and $n^{(i)}_{\rm S}(\chi)$ 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 $n_{\rm S}(z) = n_{\rm S}(\chi)\,{\rm d}\chi/{\rm d}z$. We choose an IA kernel that produces the so-called NLA model (Bridle & King 2007),

$$W^{(i)}_{\rm I}(\chi) = -A_{\rm IA}\left(\frac{1+z(\chi)}{1+z_{\rm pivot}}\right)^{\eta_{\rm IA}}\frac{C_1\rho_{\rm cr}\,\Omega_{\rm m}}{D\left(a[\chi]\right)}\,n^{(i)}_{\rm S}(\chi), \tag{16}$$

where $C_1\rho_{\rm cr} \approx 0.0134$ is constant, and $z_{\rm pivot}$ 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 $A_{\rm IA}$ and additional freedom in the redshift dependence of the IA strength via $\eta_{\rm IA}$. As there is currently no evidence for significant extra redshift evolution in our source samples (see e.g., KV450), we only consider $\eta_{\rm IA}$ in extended analyses and set $\eta_{\rm IA} = 0$ 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.

[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.

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. 2021). 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. (2020a) showed that the fiducial KV450 constraint of $A_{\rm IA} \sim 1$ 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,

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

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

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

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. 2002a) 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

$$\begin{aligned}
\xi^{(ij)}_+(\theta) &= \int_0^\infty \frac{{\rm d}\ell\ell}{2\pi} \left\{C^{(ij)}_{\epsilon\epsilon,E}(\ell) + C^{(ij)}_{\epsilon\epsilon,B}(\ell)\right\} {\rm J}_0(\ell\theta); \\
\xi^{(ij)}_-(\theta) &= \int_0^\infty \frac{{\rm d}\ell\ell}{2\pi} \left\{C^{(ij)}_{\epsilon\epsilon,E}(\ell) - C^{(ij)}_{\epsilon\epsilon,B}(\ell)\right\} {\rm J}_4(\ell\theta),
\end{aligned} \tag{19}$$

where $\rm J_\mu$ 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

$$C^{(ij)}_{E/B,l} = \frac{\pi}{\mathcal{N}_l} \int_0^\infty {\rm d}\theta\, \theta\, T(\theta) \left\{\xi^{(ij)}_+(\theta)\, g^l_+(\theta) \pm \xi^{(ij)}_-(\theta)\, g^l_-(\theta)\right\}, \tag{20}$$

where we defined the kernels

$$g^l_\pm(\theta) = \int_0^\infty {\rm d}\ell\, \ell\, S_l(\ell)\, {\rm J}_{0/4}(\ell\theta). \tag{21}$$

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 power computation in the form of a Hann window,

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

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 $x_{\rm lo} = \log\theta_{\rm lo}$ and $x_{\rm up} = \log\theta_{\rm up}$, 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 $\ell_{{\rm lo},l}$ and $\ell_{{\rm up},l}$ (for other choices, see Becker & Rozo 2016). The normalisation is then given by $\mathcal{N}_l = \ln\left(\ell_{{\rm up},i}/\ell_{{\rm lo},i}\right)$. The functions $g_\pm$ can be expressed in closed form as (Schneider et al. 2002b; van Uitert et al. 2018)

$$\begin{aligned}
g^l_+(\theta) &= \frac{1}{\theta^2}\left[\theta\ell_{{\rm up},l}\, {\rm J}_1(\theta\ell_{{\rm up},l}) - \theta\ell_{{\rm lo},l}\, {\rm J}_1(\theta\ell_{{\rm lo},l})\right]; \\
g^l_-(\theta) &= \frac{1}{\theta^2}\left[\mathcal{G}_-(\theta\ell_{{\rm up},l}) - \mathcal{G}_-(\theta\ell_{{\rm lo},l})\right],
\end{aligned} \tag{23}$$

with

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

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

$$\begin{aligned}
C^{(ij)}_{E,l} &= \frac{1}{2\mathcal{N}_l} \int_0^\infty {\rm d}\ell\, \ell \left\{W^l_{EE}(\ell)\, C^{(ij)}_{\epsilon\epsilon,E}(\ell) + W^l_{EB}(\ell)\, C^{(ij)}_{\epsilon\epsilon,B}(\ell)\right\}; \\
C^{(ij)}_{B,l} &= \frac{1}{2\mathcal{N}_l} \int_0^\infty {\rm d}\ell\, \ell \left\{W^l_{BE}(\ell)\, C^{(ij)}_{\epsilon\epsilon,E}(\ell) + W^l_{BB}(\ell)\, C^{(ij)}_{\epsilon\epsilon,B}(\ell)\right\},
\end{aligned} \tag{25}$$

with kernels given by

$$\begin{aligned}
W^l_{EE}(\ell) = W^l_{BB}(\ell) &= \int_0^\infty {\rm d}\theta\, \theta\, T(\theta) \left\{{\rm J}_0(\ell\theta)\, g^l_+(\theta) + {\rm J}_4(\ell\theta)\, g^l_-(\theta)\right\}; \\
W^l_{EB}(\ell) = W^l_{BE}(\ell) &= \int_0^\infty {\rm d}\theta\, \theta\, T(\theta) \left\{{\rm J}_0(\ell\theta)\, g^l_+(\theta) - {\rm J}_4(\ell\theta)\, g^l_-(\theta)\right\}.
\end{aligned} \tag{26}$$

The band power filters in Fourier space (Eq. (26)) and in configuration space ($T(\theta)\, g^l_\pm(\theta)$, 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

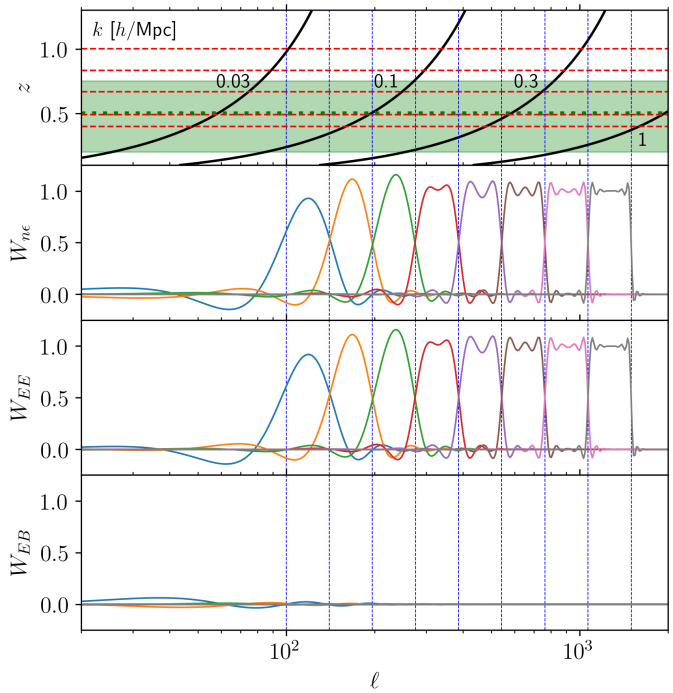

Fig. 2. Bandpower Fourier-space filters for 8 bands logarithmically spaced $\ell \in [100, 1500]$, as indicated by the vertical lines. The second to fourth panels from the top show the galaxy-galaxy lensing, cosmic shear, and *EB*-mode mixing kernels; see Eqs. (26) and (33). The *top panel* shows redshift z as a function of ℓ for a selection of wavenumbers k given in units of $h\,\mathrm{Mpc}^{-1}$. 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 $\Delta_x = 0.5$.

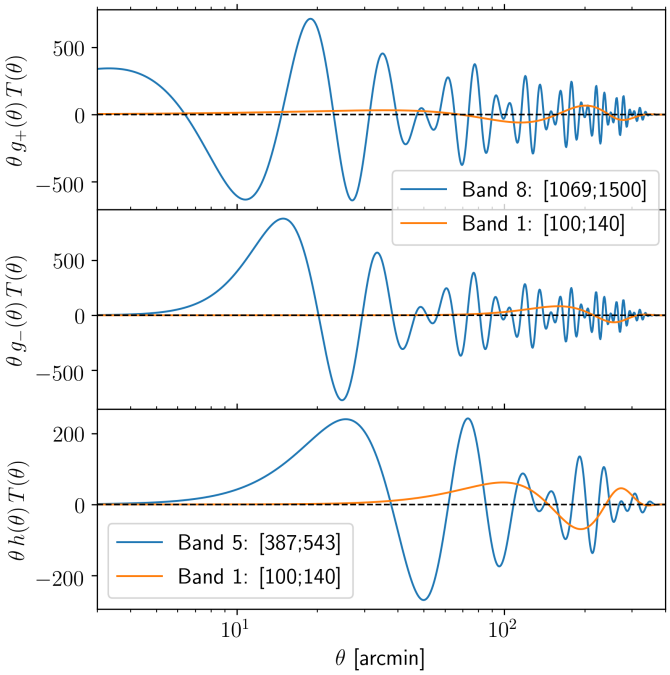

Fig. 3. Bandpower real-space filters for different bands with the angular frequency range given in square brackets. For each case we show the lowest and highest band used in the fiducial analysis setup. The default correlation function binning has been assumed, with apodisation centred on the angular 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).

a given band, our band powers have a very clean selection in Fourier space, with $W^l_{EE}(\ell)$ approaching a top hat to very good approximation. Equation (25) shows that band power *B*-modes could in principle still be generated even if $C_{\epsilon\epsilon,B} \equiv 0$, but the mode mixing via $W^l_{EB}(\ell)$ 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 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

$$C^{(ij)}_{\mathrm{n}\epsilon}(\ell) = C^{(ij)}_{\mathrm{gG}}(\ell) + C^{(ij)}_{\mathrm{gI}}(\ell) + C^{(ij)}_{\mathrm{mG}}(\ell), \tag{27}$$

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).

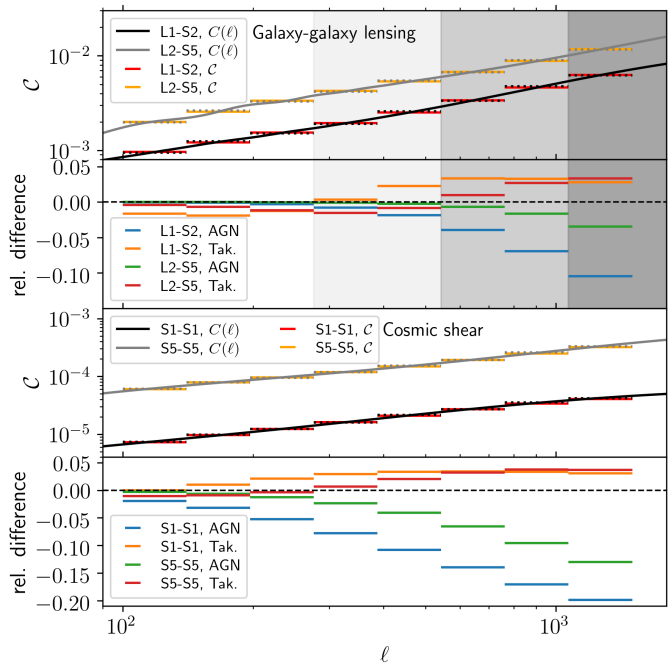

Fig. 4. Bandpowers for galaxy-galaxy lensing (*top panels*) and cosmic shear (*bottom panels*) for a selection of redshift bin combinations (indicated by 'L' and 'S' for lens and source galaxy samples, respectively; see Sect. 3 for details). Grey and black lines show the true underlying angular power spectrum; the associated dotted lines the average of these power spectra over the angular frequency band. *Bottom panels*: 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).

The line-of-sight projections read

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

for $a \in \{\rm I, G\}$, where $n_{\rm L}^{(i)}(\chi)$ 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

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

where $\alpha_{\rm mag}^{(i)}$ 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 $\alpha_{\rm mag}$, 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,

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

with

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

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

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

with the kernel

$$W_{{\rm n}\epsilon}^l(\ell) = \int_{\theta_{\rm min}}^{\theta_{\rm max}} {\rm d}\theta\, \theta\, T(\theta)\, {\rm J}_2(\ell\theta)\, h^l(\theta). \tag{33}$$

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

$$h^l(\theta) = -\frac{1}{\theta^2}\Big[\theta\ell_{{\rm up},l}{\rm J}_1(\theta\ell_{{\rm up},l}) - \theta\ell_{{\rm lo},l}{\rm J}_1(\theta\ell_{{\rm lo},l}) + 2{\rm J}_0(\theta\ell_{{\rm up},l}) - 2{\rm J}_0(\theta\ell_{{\rm lo},l})\Big]. \tag{34}$$

Figure 2 also displays the $W_{{\rm n}\epsilon}^l$ 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 $k = 1\,h\,{\rm Mpc}^{-1}$ 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 $k = 0.3\,h\,{\rm Mpc}^{-1}$ (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 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

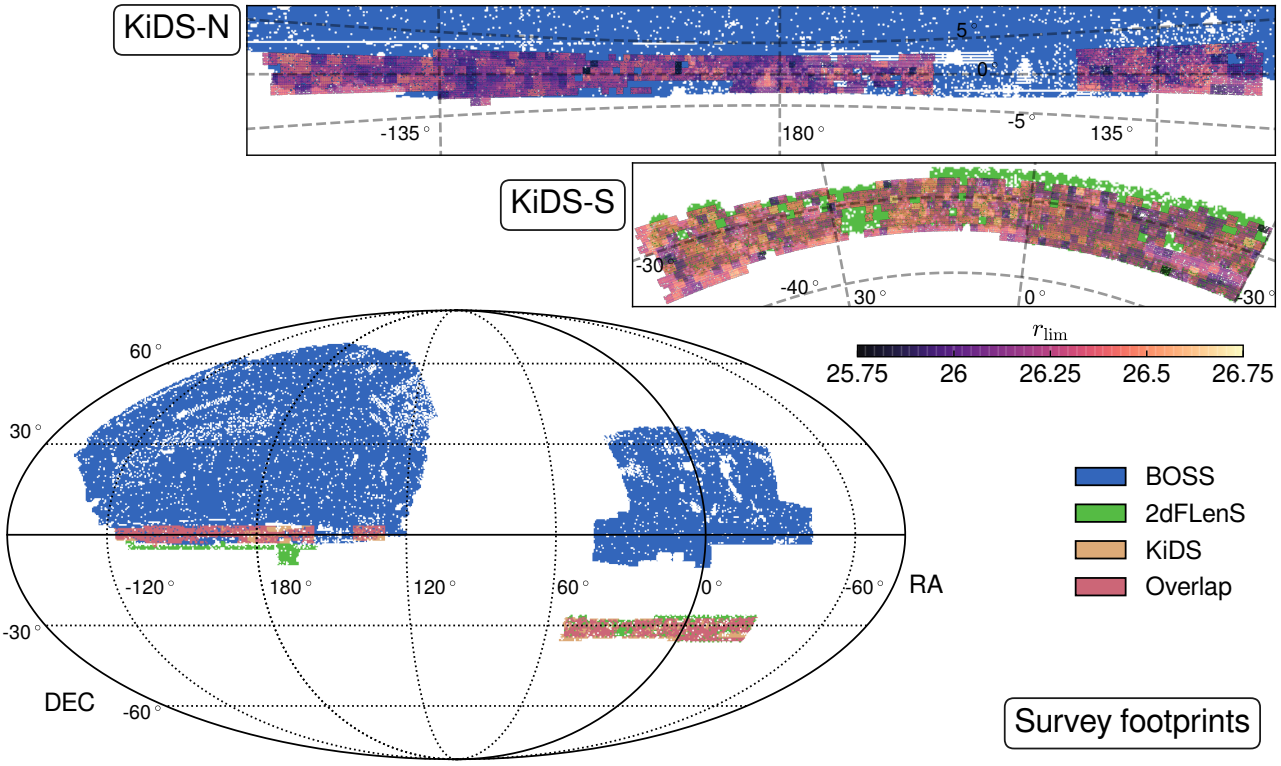

Fig. 5. *Bottom*: survey footprints of three data sets used in this analysis: BOSS (blue), 2dFLenS (green), and KiDS (orange). Overlapping regions of KiDS with either BOSS or 2dFLenS are shown in pink. *Top*: cut-outs of the KiDS areas showing the variable depth patterns of our source samples as indicated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

Table 1. Effective survey areas and sky overlap.

	KiDS-1000	BOSS	2dFLenS
Survey area	773.3 [(a)]	9329.0 [(b)]	510.8 [(c)]
KiDS overlap	–	319.5	341.9

Notes. The overlap is also illustrated in Fig. 5. All survey areas given in deg^2. Areas involving BOSS are extracted from a HEALPIX map with $N_{\rm side} = 2048$ (due to the limited size of the random catalogue used to construct the footprint); all other areas are obtained from $N_{\rm side} = 4096$ 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.

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 ∼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

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

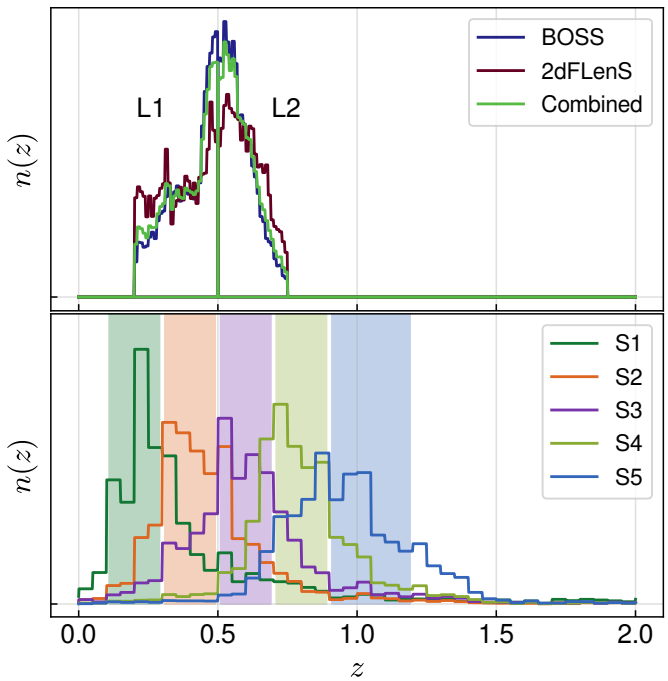

Fig. 6. Redshift distributions of the lens (*upper panel*) and source (*lower panel*) galaxy samples. Lens galaxies are split at $z = 0.5$ into two redshift bins for both the BOSS and 2dFLenS surveys. Modelling is carried out with a weighted combination of their redshift distributions (in green). Lens 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.

Tröster et al. (2020b) 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 $20\,h\,\mathrm{Mpc}^{-1}$ as a conservative minimum scale. Tröster et al. (2020a) 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. (2020a) 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.

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 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 ∼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.

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

Bin ID	z range	Mean z	$n_{\rm eff}$ [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 $n_{\rm eff}$ are for BOSS (2dFLenS). The galaxy number density $n_{\rm eff}$ 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.

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 OmegaCAM (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 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 $ugriZYJHK_s$ 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 deg^2 field of view and a mean r-band seeing of 0.7 arcsec in DR4. Improvements 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. (2021), 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 $n_{\rm S}(z)$ 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

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

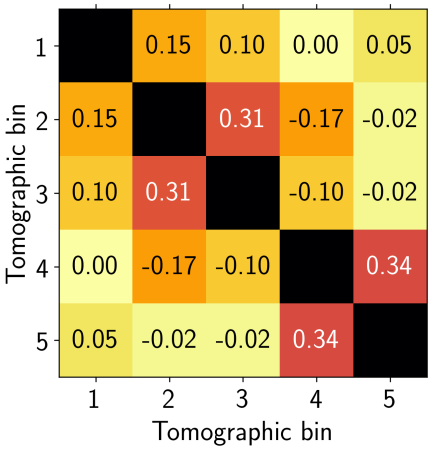

Fig. 7. Correlations between the shifts of the source redshift bins S1–5 arising from the SOM calibration. The covariance was estimated from multiple mock realisations as described in Wright et al. (2020b).

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. (2020b) applied to DR4, which means that, rather than defining spectroscopic representation via 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. 2020a).

As detailed in Wright et al. (2020b), 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 $\boldsymbol{\delta}_z$ 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. 2021 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 $\boldsymbol{\delta}_z$ prior will be centred on the mock-based estimates of the bias (Wright et al. 2020b), 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. (2020b). 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. (2021).

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. (2018a) 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. (2021) that biases proportional to PSF ellipticity ($\alpha_{\rm PSF}$) or residual PSF ellipticity after modelling ($\beta_{\rm PSF}$) introduce negligible 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 $n_{\rm eff}$, 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 $\sigma_{\rm m}$. 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 $\sigma_{\rm m}$ values for the KV450 setup.

We take a more nuanced approach to extracting $\sigma_{\rm m}$ values from the Kannawadi et al. (2019) results (see in particular their Fig. 12) than KV450 who chose a blanket value of $\sigma_{\rm m} = 0.02$ 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

$$\sigma_{\rm m} = \{0.019; 0.020; 0.017; 0.012; 0.010\} \tag{36}$$

for S1–5. As in KV450, we assume that the multiplicative bias corrections are fully correlated between the tomographic bins since the $\sigma_{\rm m}$ 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

$$\begin{aligned}\mathrm{Cov}_{\rm mult}\left[C_a^{(ij)}; C_b^{(kl)}\right] = C_a^{(ij)}\,C_b^{(kl)} \\ \times\left\{\sigma_{\rm m}^{(i)}\sigma_{\rm m}^{(k)} + \sigma_{\rm m}^{(i)}\sigma_{\rm m}^{(l)} + \sigma_{\rm m}^{(j)}\sigma_{\rm m}^{(k)} + \sigma_{\rm m}^{(j)}\sigma_{\rm m}^{(l)}\right\},\end{aligned} \tag{37}$$

where $\sigma_{\rm m}^{(i)} = 0$ 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 $n\epsilon$ for GGL. To arrive at this expression, we have assumed that the expectation of m is zero (after correction), and that $\sigma_{\rm m}^{(i)} \ll 1$ (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] 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 ∼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 ($A_{\rm c}$ 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. (2021).

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.

[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.

Two-dimensional correlation functions $\xi_{\rm gg}(s,\mu)$ are measured with a Landy-Szalay estimator (Landy & Szalay 1993) in linear bins of separation s with size $5\,h\,{\rm Mpc}^{-1}$ 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 $\Omega_{\rm m} = 0.31$. For the Landy-Szalay estimator, we adopt the random catalogues used in S17 with an oversampling factor of 50. The correlation functions include a combination 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. 2002b); 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)

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

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 $\epsilon_{{\rm t},l\to s}$ denotes the tangential ellipticity of object s measured with respect to the position of object l. The term $\mathcal{N}_{\rm rnd} := \sum_r w_r / \sum_l w_l$ 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).

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

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),

$$\hat{C}^{(ij)}_{E/B,l} = \frac{\pi}{\mathcal{N}_l} \sum_k \Delta\theta_k\, \theta_k\, T(\theta_k) \left\{ \hat{\xi}^{(ij)}_+(\theta_k)\, g^l_+(\theta_k) \pm \hat{\xi}^{(ij)}_-(\theta_k)\, g^l_-(\theta_k) \right\}, \tag{39}$$

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. 2002b; 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

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

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

We find that 300 logarithmically spaced bins between 0.5 arcmin and 300 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 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 $k \approx 1\,h\,{\rm Mpc}^{-1}$ 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 $r_{\rm lim}$ (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 and $200\,h\,{\rm Mpc}^{-1}$. 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).

[11] http://www.astro.iag.usp.br/~flask

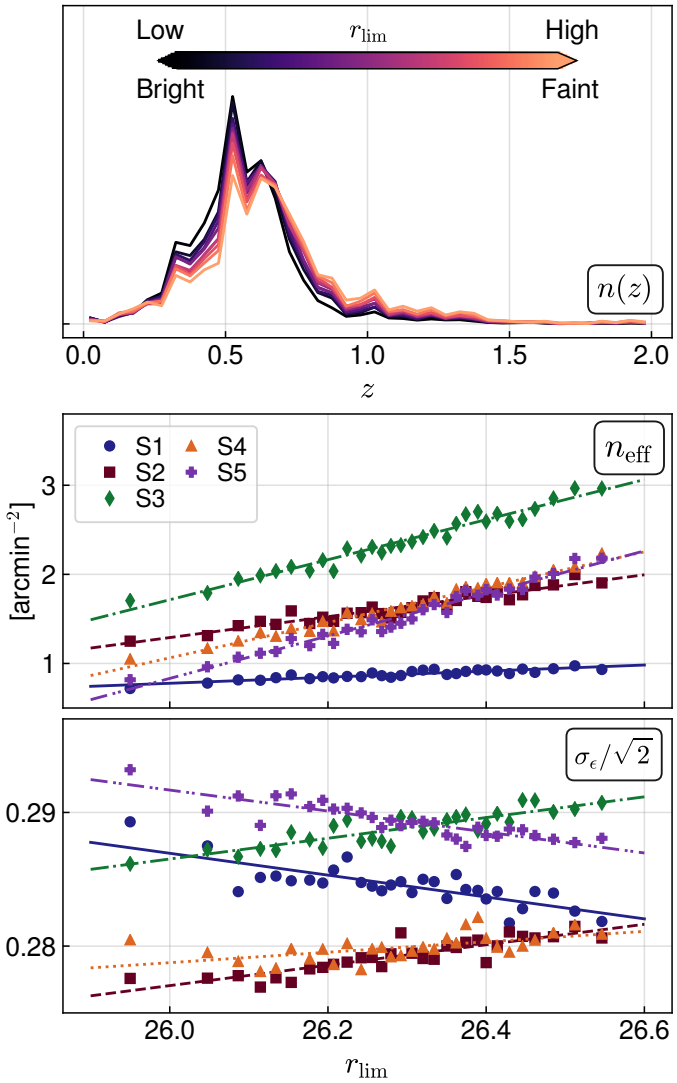

Fig. 8. Characteristic of variable depth effects measured from early KiDS-1000 data. *Top panel*: redshift distributions of source galaxies in the third tomographic bin (S3) in 10 equi-populated bins of $r_{\rm lim}$ (1σ adaptive aperture size). *Centre panel*: effective galaxy number density as a function of $r_{\rm lim}$ in the five source galaxy bins, together with linear fits. *Lower panel*: same as centre panel but for the observed ellipticity dispersion σ_ϵ.

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 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 $N_{\rm side} = 4096$, 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 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 deg^2 pointings in KV450 into 10 equi-populated bins of $r_{\rm lim}$ 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 $r_{\rm lim}$ bins in the third tomographic bin (S3). We can see that, as expected, for faint $r_{\rm lim}$ 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 $n_{\rm eff}$ and the observed ellipticity dispersion σ_ϵ on $r_{\rm lim}$. In this case galaxies from all five tomographic bins are divided into 30 equi-populated $r_{\rm lim}$ 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 $n_{\rm eff}$ 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

[12] Speedy Acquisition of Lensing and Matter Observables; the code will be made public on acceptance of the KiDS-1000 analysis papers.

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

[14] We choose to name the different mock setups after birds in order to easily distinguish between them.

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.

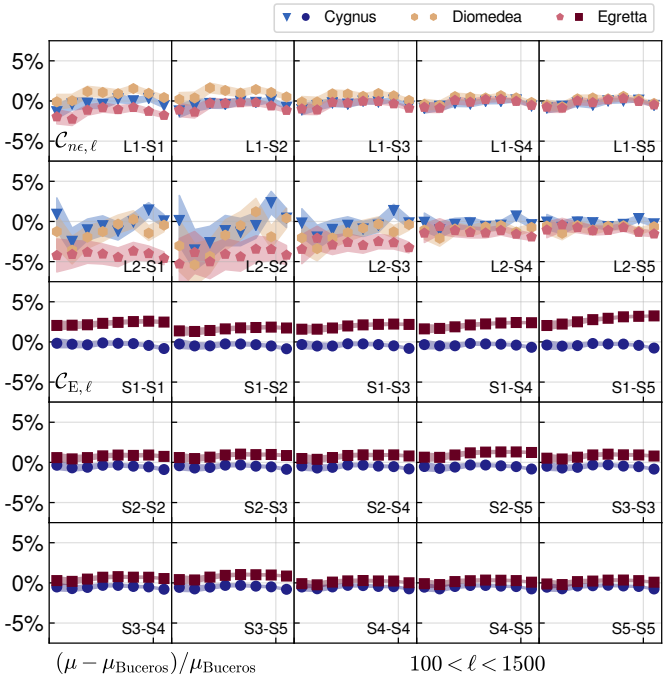

Fig. 9. Comparison of mean estimates of band power between different mock setups. Plotted is the relative deviation from the Buceros case (rectangular mask, uniform depth). Blue, yellow, and red colours correspond to the Cygnus (realistic mask, uniform depth), Diomedea (like Cygnus but using individual lens sample properties for BOSS and 2dFLenS, hence only shown for GGL), and Egretta (realistic mask and depth variations) cases, respectively. *Top two rows*: GGL signals; *bottom three*: 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.

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 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 $n_{\rm S}^{(i)}(z)$ that determine the signal amplitude, and these have a non-linear dependence on depth (cf. Fig. 8, top panel). This leads to the ∼5% negative trend in the L2-S1 and L2-S2 GGL signals, which however are not used further in our analysis.

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 $A_{\rm IA}$ and $A_{\rm bary}$, 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.

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

The model is composed of five terms of cosmological origin,

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

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 (captured 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 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. 2018b). 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 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

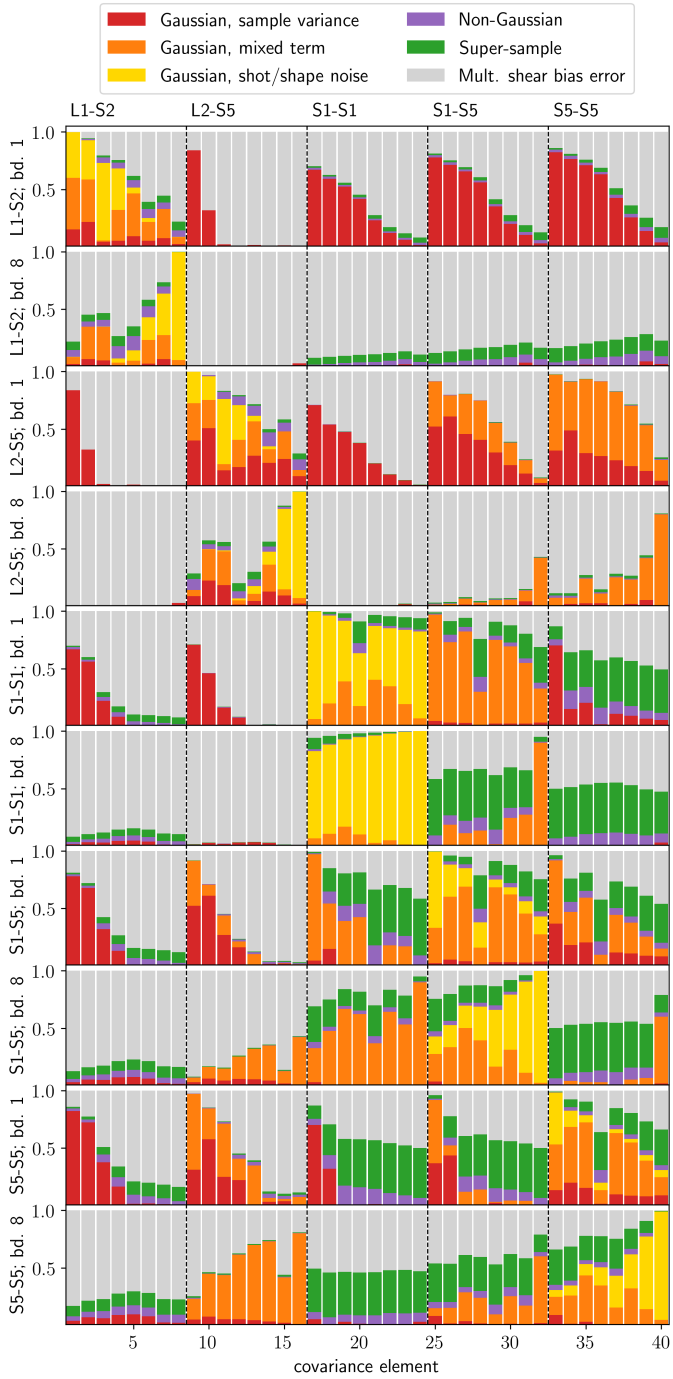

Fig. 10. Contributions to selected elements of the analytic band power covariance. Shown are the relative contributions of the absolute values of the six covariance terms: sample variance (red), mixed term (orange), and shot/shape noise (yellow) in the Gaussian covariance; non-Gaussian in-survey covariance (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.

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

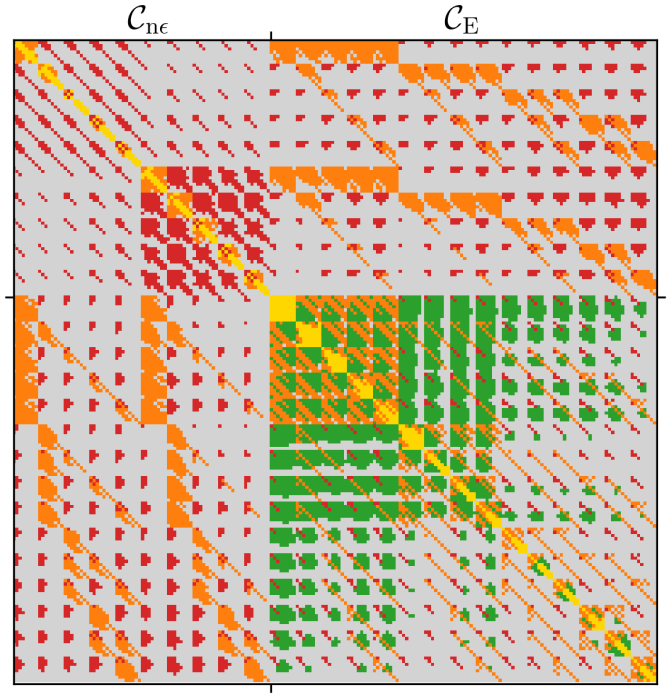

Fig. 11. Dominant contribution (absolute value) to the band power covariance. The colour coding is identical to Fig. 10, i.e. sample variance (red), mixed term (orange), and shot/shape noise (yellow) in the Gaussian covariance; non-Gaussian in-survey covariance (purple, always sub-dominant) and super-sample covariance (green); and the multiplicative shear bias uncertainty (grey). The top left block contains the covariance of the GGL signals ($C_{\mathrm{n}\epsilon}$; without cuts in tomographic bins or angular scales), the bottom right block the cosmic shear covariance (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.

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

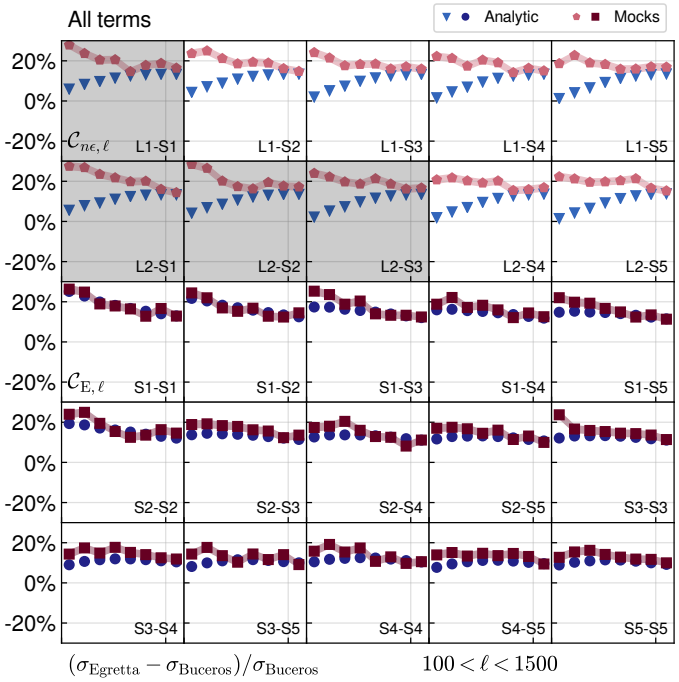

Fig. 12. Relative difference between the standard deviations of different mock setups for the covariance of the weak lensing band power signals. The relative deviation of the Egretta case (realistic footprint and depth variations) from the Buceros case (rectangular mask, uniform depth) is displayed. Red (blue) symbols show results for the mock (analytic) covariance. *Top two rows*: GGL signals; *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.

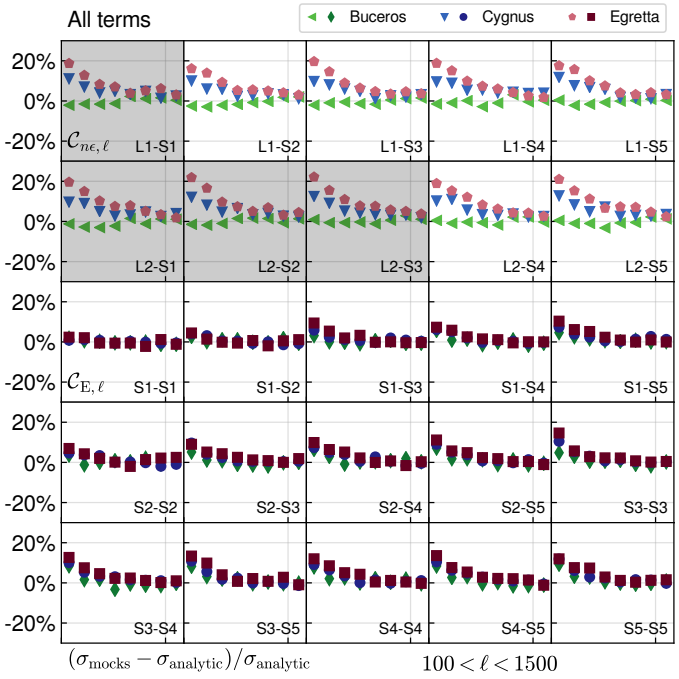

Fig. 13. Relative difference between the square root of the diagonals of the mock and analytic covariances for the weak lensing band power signals. *Top two rows*: GGL signals; *bottom three*: cosmic shear signals, with bin combinations indicated in the panels. Three cases are shown: spatially uniform 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.

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. 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.

[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.

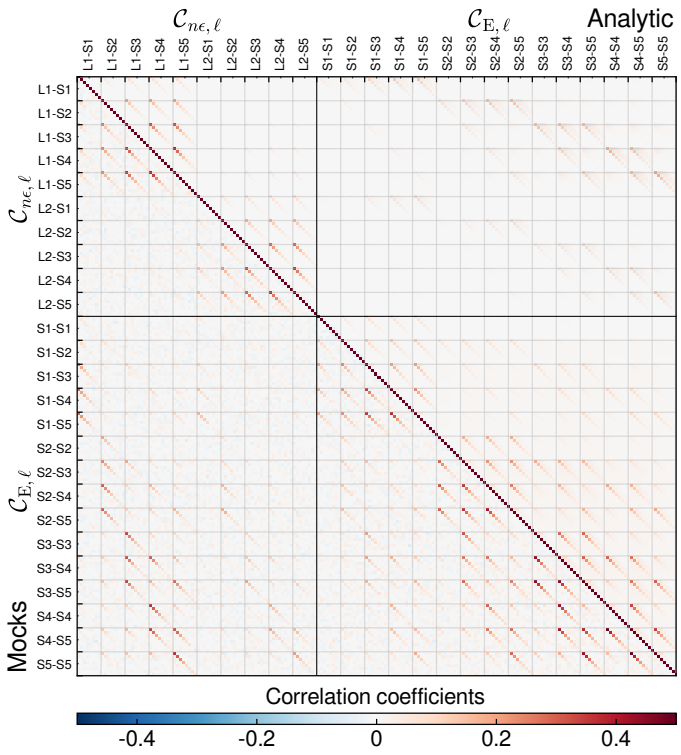

Fig. 14. Comparison of correlation coefficients of the weak lensing band power covariance between mocks (lower left) and analytic covariance (upper right) in the most realistic setup (Egretta). The top left block is for GGL; the bottom right block for cosmic shear. Galaxy sample combinations are indicated on the axes.

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.

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 $i \in \{\mathrm{A}, \mathrm{B}\}$, 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, $\mathrm{Cov} = \mathsf{L}\mathsf{L}^\intercal$.

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

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

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 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 $S_8 = \sigma_8 \, (\Omega_\mathrm{m}/0.3)^{1/2}$, approximately measuring the lateral width of the posterior across the aforementioned degeneracy. We note however that the exact direction of the Ω_m–σ_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. 2018a), respectively.

[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

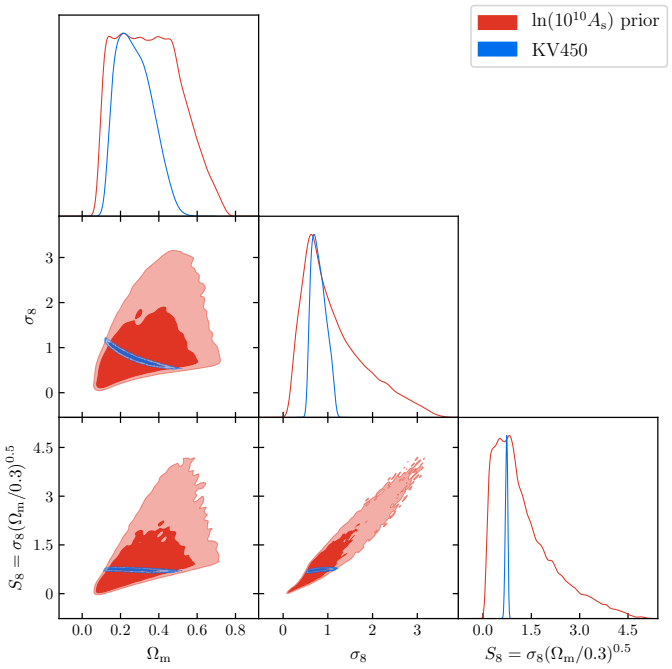

Fig. 15. Prior volume resulting from using wide top-hat priors on the sampling parameters $\{\ln A_s, \omega_c, \omega_b, n_s, h\}$ (in red), shown together with the KV450 posterior (in blue). Dark (light) shades correspond to the two-dimensional 1σ (2σ) credible regions. The prior does not impact on marginal S_8 constraints, but is informative in the Ω_m–S_8 and Ω_m–σ_8 planes.

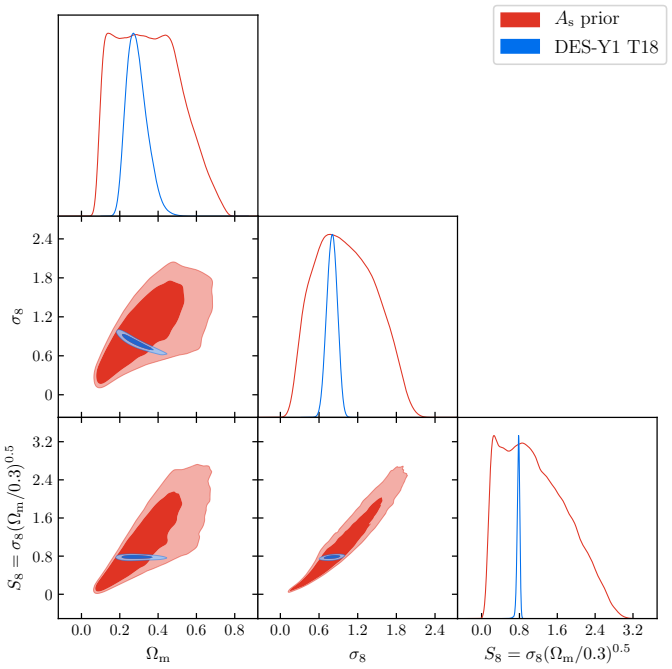

Fig. 16. Same as Fig. 15, but now showing the prior volume using the sampling parameter A_s with a wide top-hat prior, alongside the DES Year 1 cosmic shear posterior. It should be noted that the axis range for the σ_8 and S_8 parameters is smaller than in Fig. 15.

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 Ω_m–σ_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.

[20] By probability mass, we refer to the probability density function (PDF) integrated over certain regions of the PDF's domain.

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 encompasses 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. (2020a) 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,

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

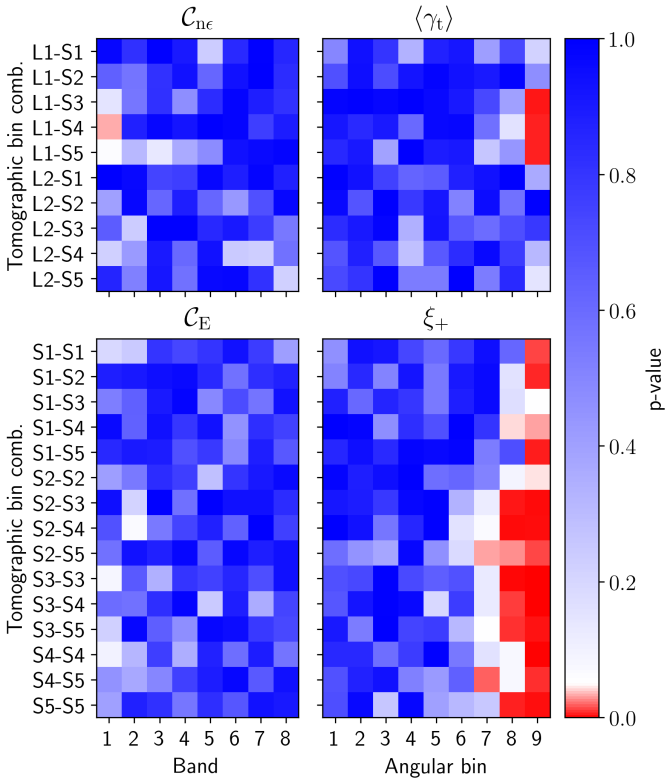

Fig. 17. *p*-values for a KS-test of Gaussianity of the marginal likelihood distributions. Values above the typical threshold of 0.05 are blue; values below are red. *Left panels*: band powers for cosmic shear (*bottom*) and galaxy-galaxy lensing (*top*) for various tomographic bin combinations and angular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

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 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. (2020a) 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

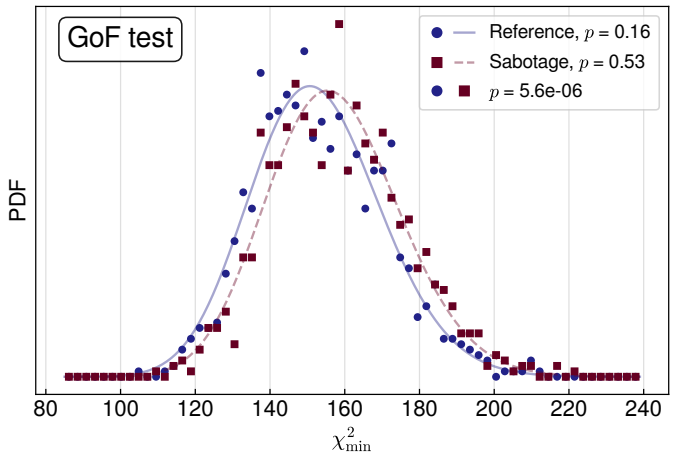

Fig. 18. Goodness of fit sampling distributions. Shown are histograms of minimum χ^2-values obtained from 1000 mock realisations of data vectors analysed with our cosmic shear and GGL analysis pipeline. Blue points correspond to our baseline setup; red points to the result using a sabotaged data vector 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 ($p < 10^{-5}$).

it follows a χ^2 distribution with $k_{\rm d.o.f.} = N_{\rm d} - N_\Theta$ 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 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 $k_{\rm d.o.f.,eff} := N_{\rm d} - N_{\Theta,\rm eff}$. 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 have seen applications in the astrophysical literature:

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

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

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

where $\Pr(\boldsymbol{\Theta}|\boldsymbol{d})$ 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 $N_{\Theta,\rm eff}$ than Eq. (46), it does not perform favourably in our setup.

Figure 19 shows the performance of the estimators in relation to $N_{\rm d} - k_{\rm d.o.f.,eff}$ as inferred from the $\chi^2_{\rm min}$ sampling distribution (cf. Fig. 18). Neither estimator is a good match per se. Since $N_{\Theta,\rm var}$ 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, $N_{\Theta,\rm eff} := (N_{\Theta,\rm like} + N_{\Theta,\rm post})/2$.

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

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

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 $k_{\rm d.o.f.,eff}$ degrees of freedom to the $\chi^2_{\rm min}$ 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

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