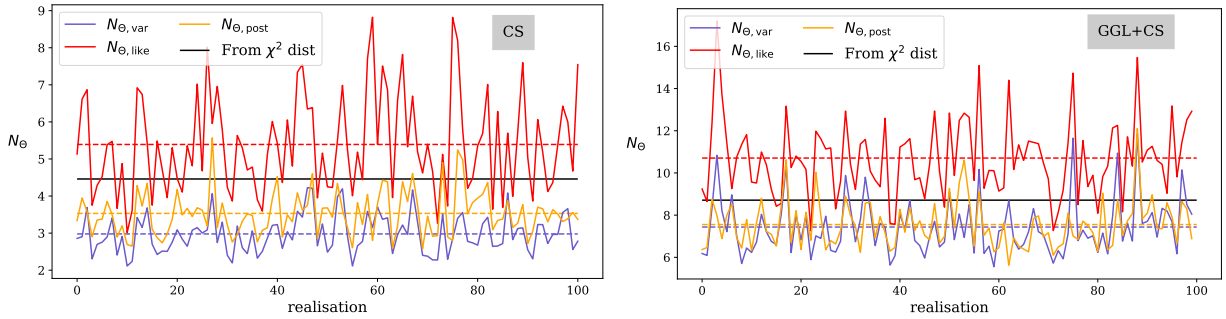

Fig. 19. Estimates of the effective number of parameters for 100 realisations of a cosmic shear (*left*) and a joint cosmic shear+GGL (*right*) data vector. The estimators defined in Eqs. (44)–(46) are shown in blue, red, and orange, respectively. The dashed lines mark the average over the realisations. The black line marks the effective number of parameters inferred from the fit of a χ^2-distribution to a histogram of χ^2_{min} (cf. Eq. (47)).

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

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

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

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

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

6.4. Reporting parameter constraints

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

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

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

$$\boldsymbol{\Theta}_{\mathrm{MAP}} = \mathrm{argmax}_{\boldsymbol{\Theta}} \Pr(\boldsymbol{\Theta}|\boldsymbol{d}), \tag{48}$$

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

The decision-theoretical optimal uncertainty estimate for the MAP is the credible region $\mathcal{H}_\alpha$ defined by the highest posterior density (Robert 2001; see also Price et al. 2021 for another recent application in cosmology)

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

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

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

[22] Pathological examples with arbitrarily low marginal posterior density at the position of the maximum of the joint posterior are straightforward to construct.

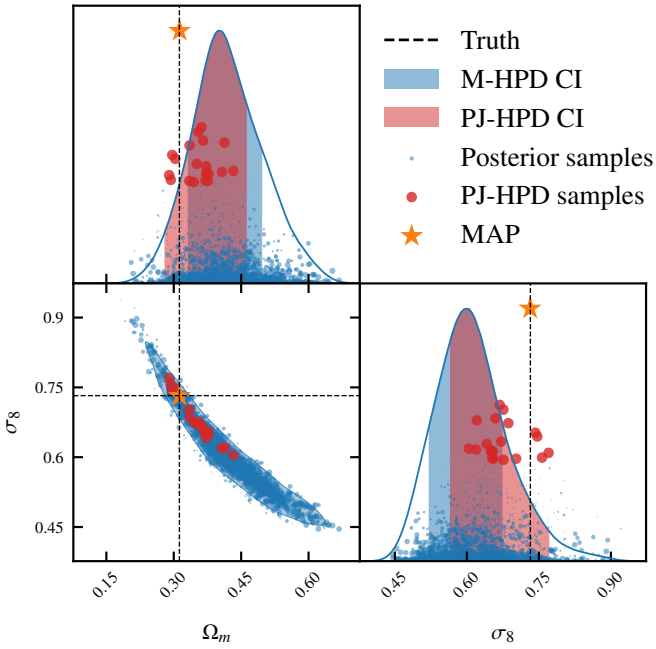

Fig. 20. Illustration of 1σ credible intervals (CI) used to report parameter constraints, for a subset of two parameters for an exemplary mock joint analysis of GGL and cosmic shear with a noiseless data vector. The standard marginal highest posterior density (M-HPD) CIs are shown in blue. The marginal posterior mode is shifted with respect to the true input parameter values (black dashed lines), whereas the maximum a posteriori (MAP) estimate (orange star) tracks the truth well. The CI we associate with the MAP is constructed via the PJ-HPD method (red), using the samples with the highest posterior density (red points). Other posterior samples are shown as blue points, with point size proportional to their posterior mass. *Top* and *right panels*: height of the points and of the MAP indicate their posterior density with arbitrary scaling.

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

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

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

such that

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

where $\Pr(\Theta_i|\boldsymbol{d})$ is the posterior of Θ_i marginalised over all other parameters of the model. In general, this does not correspond to the region of the highest marginal posterior. The PJ-HPD estimate is guaranteed to include the MAP (since by definition MAP lies within the highest posterior density region) and readily allows for the comparison of the credible intervals between analyses with a different dimension of parameter space. It is intuitive in that the PJ-HPD reduces to the standard M-HPD credible interval in one dimension. Moreover, in the case of a multivariate Gaussian posterior, the PJ-HPD credible interval coincides with the marginal 1σ interval.

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

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

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

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

Finally, we note that posterior sampling techniques have an inherent limit on the accuracy of parameter constraints,

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

dependent on the number of samples drawn. For settings that our computing resources can deliver in reasonable time (always yielding well in excess of 10^4 posterior samples), we find this limit to be $\sim 0.1\sigma$ in the key parameter S_8, corresponding to the scatter of best-fit values between repeat runs of otherwise identical setups. This floor sets a stopping criterion for validation, as well as for the mitigation of systematic effects in our data (as reported in Giblin et al. 2021).

7. Validation of the likelihood analysis pipeline

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

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

In Fig. 21 we show one- and two-dimensional marginal posteriors of a subset of the 20 sampling parameters that are significantly constrained, alongside projections of the MAP. We also include Ω_m and σ_8 as derived parameters. The MAP recovers the input values of all parameters shown in the figure to better than 1.7%. The combination of clustering and weak lensing signals partially lifts the typical degeneracy between Ω_m and σ_8 for weak lensing-only constraints, through the small clustering errors on Ω_m. This benefits the marginal posterior of the key parameter S_8 as well since in our analysis setup it is still significantly correlated with Ω_m. The joint analysis also leads to visibly tighter and more symmetric marginal posteriors for the linear bias parameters, whereas improvements on the constraints on intrinsic alignments, baryon feedback, and the calibration of the redshift distributions remain marginal.

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

The 1σ M-HPD credible interval for cosmic shear alone is 20% smaller than the KV450 constraint, while the fiducial joint analysis halves the statistical uncertainty on S_8 with respect to KV450 (which corresponds to 29% smaller errors than the corresponding BOSS+KV450 constraints by Tröster et al. 2020a). The marginal 1σ credible intervals shrink by 10% (18%) for Ω_m over BOSS+KV450 (BOSS alone), and by 31% (40%) for A_IA over BOSS+KV450 (KV450 alone). The precision on linear and non-linear galaxy bias parameters increases by roughly 30% on most parameters with respect to BOSS only, but most of the added constraining power was already achieved in the Tröster et al. (2020a) analysis, and non-linear bias constraints are still too weak for informative conclusions on galaxy physics.

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

Naively, one might expect the KiDS-only parameter errors to decrease by a factor $\sqrt{2}$ because the KiDS-1000 survey area has doubled with respect to earlier KiDS analyses. However, this does not take into account the highly non-linear parameter dependencies, updated priors, or covariance contributions that do not scale inversely with area, in particular the multiplicative shear bias uncertainty. Most importantly in the cosmic shear case though, for our choices of two-point statistic and scales included in the analysis S_8 is not an optimal summary parameter of the constraining power of the data, as significant correlations with Ω_m remain (see Fig. 23). While a power-law index of 0.5 in the S_8 definition was close to optimal for the $\xi_\pm$ data analysed in KV450, we find that larger values of $\sim$0.6 for the index capture the degeneracy between Ω_m and σ_8 better for band powers (this is explored further in Asgari et al. 2021a). The gain in constraining power for an adjusted definition of S_8 is larger and consistent with the naive $\sqrt{2}$ expectation, indicating that the systematic error treatment and the analysis choices have actually evolved in tandem with the statistical power of KiDS-1000.

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

Since we exclude bin combinations with strong overlap of sources and lenses, there is also no self-calibration of intrinsic alignments so that marginal A_IA constraints are the same for the joint analysis and cosmic shear alone; see Fig. 21. On the real data the addition of GGL to cosmic shear and clustering does lead to a tightening of the marginal A_IA posterior by 17%, but the marginal posterior maximum also shifts by 12% (Heymans et al. 2021). Both real and mock analysis agree that GGL does not significant constraining power to the galaxy bias parameters either.

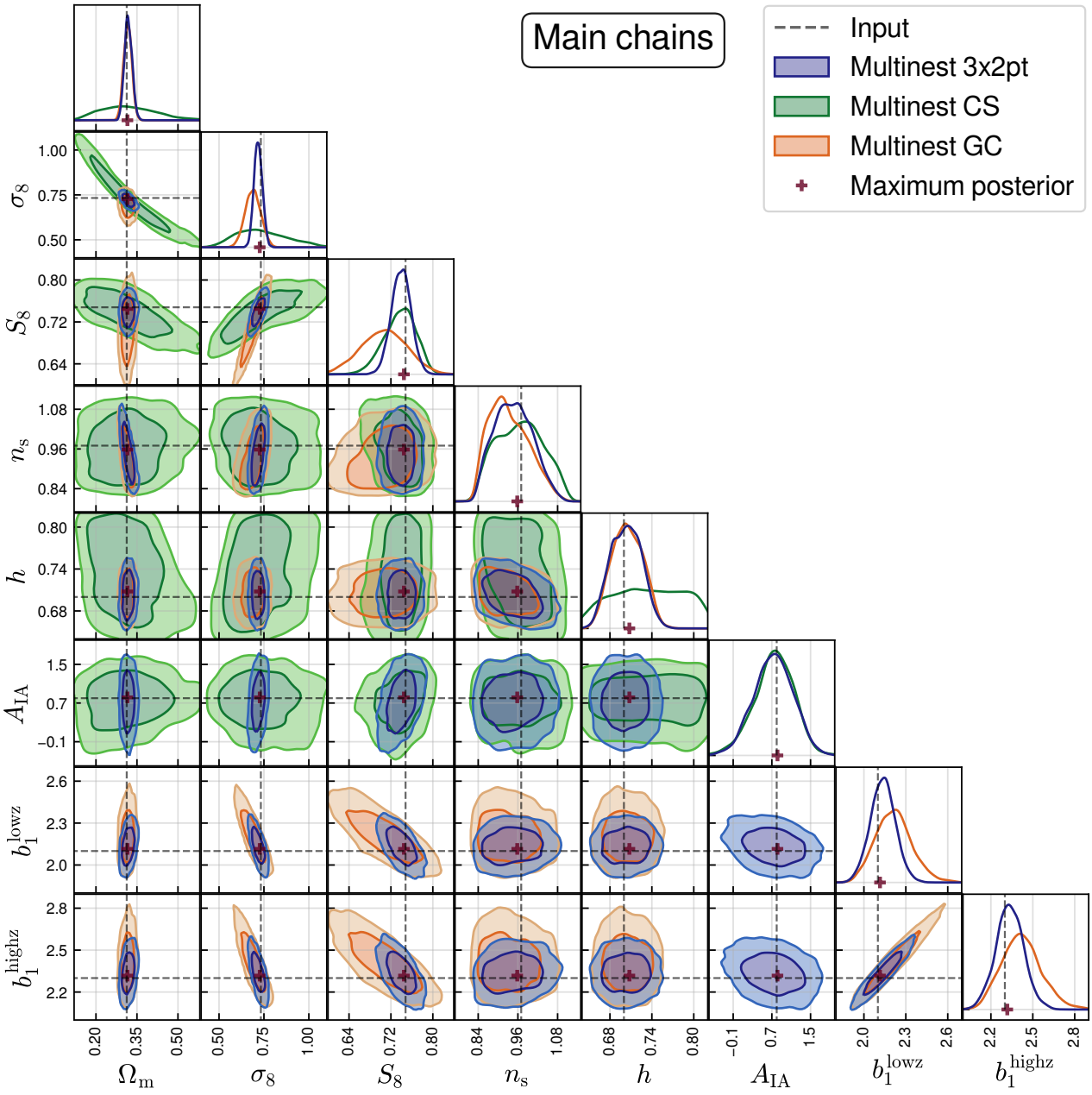

Fig. 21. Marginal posterior distributions for a selection of parameters that are significantly constrained by the data, including the derived parameters Ω_m and σ_8. Shown are one-dimensional marginals along the diagonal and two-dimensional marginal 1σ and 2σ credible regions elsewhere. Results are for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

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

The S_8 constraints are also stable under changes from an uncorrelated prior on the source redshift distribution shifts δ_z, to the correlated prior in Fig. 7, as well as under the choice of

[24] For these analyses we kept the δz parameter fixed at 0 to speed up the runs and did not include the uncertainty due to multiplicative shear bias calibration to highlight the differences in the cosmological covariance contributions.

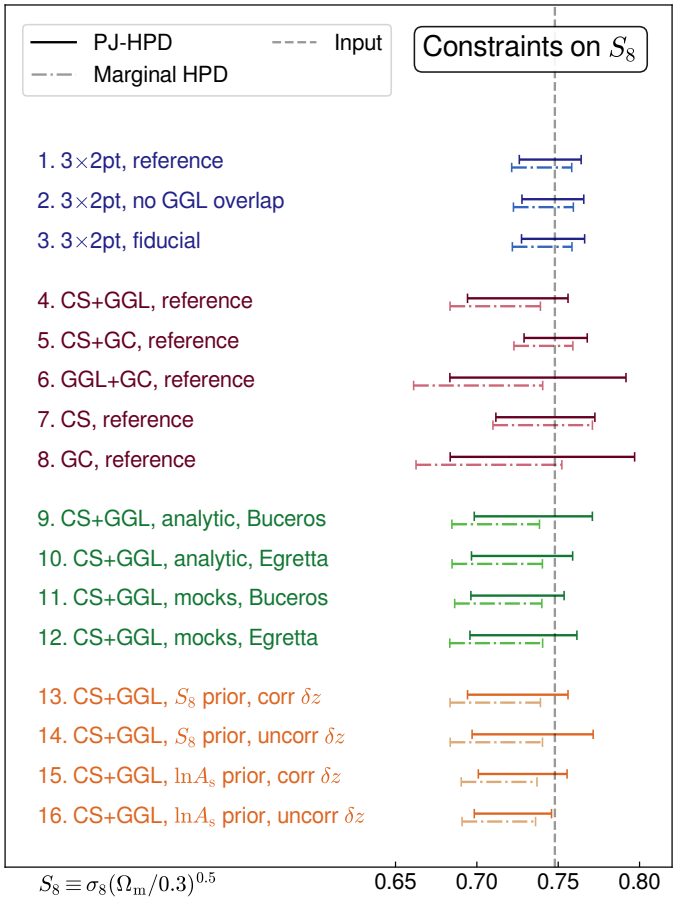

Fig. 22. S_8 constraints (1σ) for different compositions of the GGL data vector in the joint analysis (blue), probe combinations (red), different covariance models (green), and different prior assumptions (orange). Standard marginal (M-)HPD credible intervals are shown as dot-dashed lines; PJ-HPD intervals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

Table 5. Validation tests and resulting S_8 marginal credible intervals.

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

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

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

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

We note that, for an assessment of consistency with *Planck* constraints, the *Planck* likelihood will need to be sampled in S_8 as well. The impact of the choice of sampling parameter and the displacement of one-dimensional marginal constraints away from the underlying best-fit value highlight that discrepancies between probes should be interpreted with care if quantified solely in the marginal distributions, as is widespread practice. Tension is assessed most meaningfully in the shared parameter space of the full posterior distributions, but this can be challenging in practice. Therefore, we still assess tension via the marginal distributions of S_8, but only for probe combinations for which we find projection effects to be negligible. A detailed discussion of tension assessment is provided in Heymans et al. (2021). Bayesian approaches to performing consistency tests in large parameter spaces exist (Köhlinger et al. 2019) and have been applied to the internal consistency of KiDS-1000 cosmic shear (Asgari et al. 2021a). We advocate the use of a summary measure of tension that minimises the sensitivity to prior choices (see e.g., Handley & Lemos 2019b; Lemos et al. 2020).

We find no measurable difference between M-HPD and PJ-HPD S_8 credible intervals for *Planck* (as expected since the

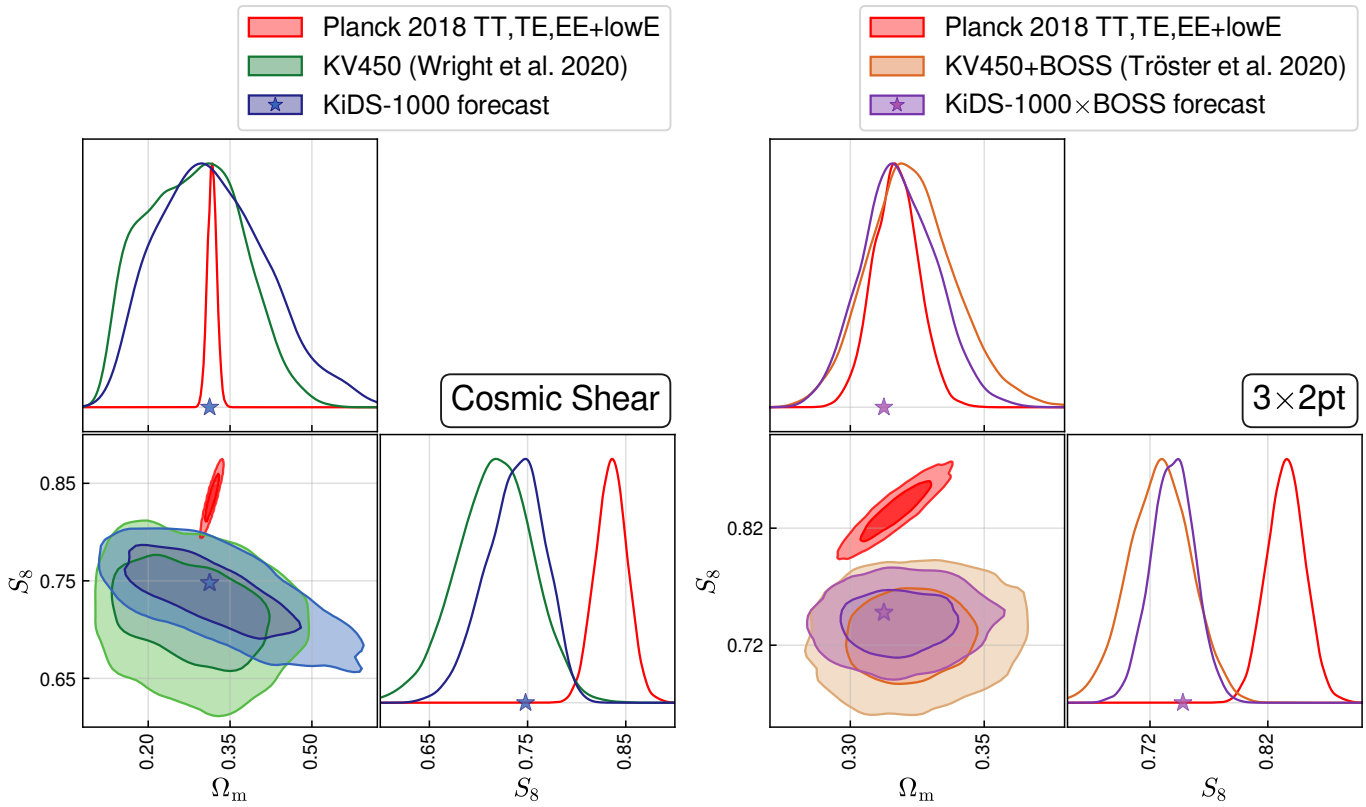

Fig. 23. Predicted KiDS-1000 constraints (reference setup, IDs 1 and 7) on the key cosmological parameters Ω_m and S_8 in relation to previous KiDS results (Wright et al. 2020a; Tröster et al. 2020a) and the Planck Collaboration VI (2020) primary CMB constraints, as indicated in the legends. *Left panel*: results for cosmic shear only; *right panel*: those for the joint analysis of BOSS clustering, galaxy-galaxy lensing between BOSS/2dFLenS and KiDS-1000, and cosmic shear. The star in each panel marks the fiducial cosmology assumed for the forecast. Axis scales differ between the two panels.

cosmological posterior is close to Gaussian), and small shifts for KV450 in line with the results of Fig. 22, so that the previously reported S_8 tension between these probes remains valid. The marginal S_8 posterior for KiDS-1000 cosmic shear only and for the joint analysis of clustering and weak lensing peaks close to the input and is well approximated by a Gaussian (cf. Fig. 23), so direct comparison with *Planck* is possible for this case as well.

8. Summary and conclusions

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

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

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

The main source of uncertainty in the modelling of weak lensing, however, remains the intrinsic alignment of galaxies (Sect. 2.4). Direct observations of the effect in galaxy samples typically used for weak lensing do not yet exist, and our physical understanding of alignment mechanisms is still too poor to create predictive ab-initio simulations. In the absence of clear guidance

we are thus required to strike a balance between a simple model that risks not capturing the complexity in the data and a flexible model that risks providing catch-all parameters in the likelihood analysis among which other residual systematics, notably those related to the source redshift distributions, can hide.

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

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

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

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

With the fiducial KiDS-1000 analysis setup, we can expect cosmic shear-only S_8 constraints that are 20% tighter than KV450, and joint clustering and weak lensing constraints that improve in S_8 by 29% over the previous KiDS+BOSS analysis (Sect. 7). These changes undersell the true gain in statistical power somewhat because, as opposed to the correlation function-based analysis in KV450, S_8 does not precisely capture the direction transverse to the typical weak lensing Ω_m–σ_8 degeneracy any more.

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

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

Acknowledgements. We would like to thank M. Jarvis for his continued excellent support and maintenance of TreeCorr. We are also grateful to E. Tittley who went out of his way to preserve data in peril. We thank our referee for constructive comments on the manuscript. CAL, MA, TT, CH, and BG acknowledge support from the European Research Council under grant number 647112. TT is also supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 797794. CH and FK are supported by the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research. HHi is supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1). HHi, AHW, JLvdB, and AD are supported by an ERC Consolidator Grant (No. 770935). AGS acknowledges support by the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. MB is supported by the Polish Ministry of Science and Higher Education through grant DIR/WK/2018/12, and by the Polish National Science Center through grant no. 2018/30/E/ST9/00698. MC acknowledges support by the Spanish Ministry of Science MINECO under grant PGC2018-102021. HHo and

[25] https://sci.esa.int/euclid
[26] Legacy Survey of Space and Time; www.lsst.org
[27] Dark Energy Spectroscopic Instrument; www.desi.lbl.gov
[28] www.4most.eu
[29] Subaru Prime Focus Spectrograph; https://pfs.ipmu.jp

AK acknowledge support from Vici grant 639.043.512, financed by the Netherlands Organisation for Scientific Research (NWO). KK acknowledges support by the Alexander von Humboldt Foundation. HYS acknowledges the support from NSFC of China under grant 11973070, the Shanghai Committee of Science and Technology grant No.19ZR1466600 and Key Research Program of Frontier Sciences, CAS, Grant No. ZDBS-LY-7013. MvWK acknowledges the support by the Science and Technology Facilities Council. We are indebted to the staff at ESO-Garching and ESO-Paranal for managing the observations at VST and VISTA that yielded the data presented here. Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018, 179.A-2004, 298.A-5015, and on data products produced by the KiDS consortium. The 2dFLenS survey is based on data acquired through the Australian Astronomical Observatory, under program A/2014B/008. It would not have been possible without the dedicated work of the staff of the AAO in the development and support of the 2dF-AAOmega system, and the running of the AAT. The BOSS-related results in this paper have been made possible thanks to SDSS-III. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. Author Contributions: All authors contributed to the development and writing of this paper. The authorship list is given in three groups: the lead authors (BJ, CAL, MA, TT, CH), followed by two alphabetical groups. The first alphabetical group includes those who are key contributors to both the scientific analysis and the data products.The second group covers those who have either made a significant contribution to the data products or to the scientific analysis.

References

Abbott, T., Abdalla, F. B., Allam, S., et al. 2016, Phys. Rev. D, 94, 022001
Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys. Rev. D, 98, 043526
Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Alcock, C., & Paczynski, B. 1979, Nature, 281, 358
Amara, A., & Réfrégier, A. 2007, MNRAS, 381, 1018
Asgari, M., & Schneider, P. 2015, A&A, 578, A50
Asgari, M., Heymans, C., Hildebrandt, H., et al. 2019, A&A, 624, A134
Asgari, M., Tröster, T., Heymans, C., et al. 2020, A&A, 634, A127
Asgari, M., Lin, C. A., Joachimi, B., et al. 2021a, A&A, 645, A104
Asgari, M., Friswell, I., Yoon, M., et al. 2021b, MNRAS, 501, 3003
Barreira, A., Krause, E., & Schmidt, F. 2018a, JCAP, 2018, 053
Barreira, A., Krause, E., & Schmidt, F. 2018b, JCAP, 2018, 015
Bartelmann, M., & Schneider, P. 2001, Phys. Rep., 340, 291
Becker, M. R., & Rozo, E. 2016, MNRAS, 457, 304
Becker, M. R., Troxel, M. A., MacCrann, N., et al. 2016, Phys. Rev. D, 94, 022002
Benítez, N. 2000, ApJ, 536, 571
Bernardeau, F., van Waerbeke, L., & Mellier, Y. 1997, A&A, 322, 1
Bernstein, G. M. 2009, ApJ, 695, 652
Blake, C., Brough, S., Colless, M., et al. 2011, MNRAS, 415, 2876
Blake, C., Amon, A., Childress, M., et al. 2016, MNRAS, 462, 4240
Blake, C., Amon, A., Asgari, M., et al. 2020, A&A, 642, A158
Blazek, J. A., MacCrann, N., Troxel, M. A., & Fang, X. 2019, Phys. Rev. D, 100, 103506
Bose, B., Winther, H. A., Pourtsidou, A., et al. 2020, JCAP, 2020, 001
Bridle, S., & King, L. 2007, New J. Phys., 9, 444
Cacciato, M., Lahav, O., van den Bosch, F. C., Hoekstra, H., & Dekel, A. 2012, MNRAS, 426, 566
Carretero, J., Castander, F. J., Gaztañaga, E., Crocce, M., & Fosalba, P. 2015, MNRAS, 447, 646
Carron, J. 2013, A&A, 551, A88
Chan, K. C., Scoccimarro, R., & Sheth, R. K. 2012, Phys. Rev. D, 85, 083509
Chisari, N. E., Mead, A. J., Joudaki, S., et al. 2019, Open J. Astrophys., 2, 4
Colas, T., d'Amico, G., Senatore, L., Zhang, P., & Beutler, F. 2020, JCAP, 2020, 001
Crocce, M., & Scoccimarro, R. 2006, Phys. Rev. D, 73, 063519
Crocce, M., Scoccimarro, R., & Bernardeau, F. 2012, MNRAS, 427, 2537
Crocce, M., Castander, F. J., Gaztañaga, E., Fosalba, P., & Carretero, J. 2015, MNRAS, 453, 1513
d'Amico, G., Gleyzes, J., Kokron, N., et al. 2020, JCAP, 2020, 005
de Putter, R., Wagner, C., Mena, O., Verde, L., & Percival, W. J. 2012, JCAP, 2012, 019
Dekel, A., & Lahav, O. 1999, ApJ, 520, 24
Deshpande, A. C., Kitching, T. D., Cardone, V. F., et al. 2020, A&A, 636, A95
DESI Collaboration (Aghamousa, A., et al.) 2016, ArXiv e-prints [arXiv:1611.00036]
Desjacques, V., Jeong, D., & Schmidt, F. 2018, Phys. Rep., 733, 1
Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64
Duncan, C. A. J., Joachimi, B., Heavens, A. F., Heymans, C., & Hildebrandt, H. 2014, MNRAS, 437, 2471
Dvornik, A., Hoekstra, H., Kuijken, K., et al. 2018, MNRAS, 479, 1240
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Efstathiou, G., & Lemos, P. 2018, MNRAS, 476, 151
Eggemeier, A., Scoccimarro, R., Crocce, M., Pezzotta, A., & Sánchez, A. G. 2020, Phys. Rev. D, 102, 103530
Eifler, T., Krause, E., Dodelson, S., et al. 2015, MNRAS, 454, 2451
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Erben, T., Schirmer, M., Dietrich, J. P., et al. 2005, Astron. Nachr., 326, 432
Euclid Collaboration (Knabenhans, M., et al.) 2019, MNRAS, 484, 5509
Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23
Fenech Conti, I., Herbonnet, R., Hoekstra, H., et al. 2017, MNRAS, 467, 1627
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, Open J. Astrophys., 2, 10
Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2021, MNRAS, 501, 2983
Fosalba, P., Crocce, M., Gaztañaga, E., & Castander, F. J. 2015a, MNRAS, 448, 2987
Fosalba, P., Gaztañaga, E., Castander, F. J., & Crocce, M. 2015b, MNRAS, 447, 1319
Gelman, A., Meng, X. L., & Stern, H. 1992, Tech. Report #355, Dept. of Statistics, University of Chicago
Gelman, A., Meng, X.-L., & Stern, H. 1996, Stat. Sin., 6, 733
Giblin, B., Heymans, C., Asgari, M., et al. 2021, A&A, 645, A105
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Griffith, R. L., Cooper, M. C., Newman, J. A., et al. 2012, ApJS, 200, 9
Hamilton, A. J. S., Kumar, P., Lu, E., & Matthews, A. 1991, ApJ, 374, L1
Hamimeche, S., & Lewis, A. 2008, Phys. Rev. D, 77, 103013
Handley, W., & Lemos, P. 2019a, Phys. Rev. D, 100, 023512
Handley, W., & Lemos, P. 2019b, Phys. Rev. D, 100, 043504
Harnois-Déraps, J., Amon, A., Choi, A., et al. 2018, MNRAS, 481, 1337
Harnois-Déraps, J., Giblin, B., & Joachimi, B. 2019, A&A, 631, A160
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Hartlap, J., Hilbert, S., Schneider, P., & Hildebrandt, H. 2011, A&A, 528, A51
Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2014, ApJ, 780, 111
Heydenreich, S., Schneider, P., Hildebrandt, H., et al. 2020, A&A, 634, A104
Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
Heymans, C., Grocutt, E., Heavens, A., et al. 2013, MNRAS, 432, 2433
Heymans, C., Tröster, T., Asgari, M., et al. 2021, A&A, in press, https://doi.org/10.1051/0004-6361/202039063
Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
Hilbert, S., Hartlap, J., & Schneider, P. 2011, A&A, 536, A85
Hildebrandt, H. 2016, MNRAS, 455, 3943
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, in press, https://doi.org/10.1051/0004-6361/202039018
Hirata, C. M., & Seljak, U. 2004, Phys. Rev. D, 70, 063526
Hoffmann, K., Bel, J., Gaztañaga, E., et al. 2015, MNRAS, 447, 1724
Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, JCAP, 2012, 027
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Huang, H.-J., Eifler, T., Mandelbaum, R., & Dodelson, S. 2019, MNRAS, 488, 1652
Huff, E., & Mandelbaum, R. 2017, ArXiv e-prints [arXiv:1702.02600]
Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, JCAP, 2020, 042
Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338
Joachimi, B., & Bridle, S. L. 2010, A&A, 523, A1
Joachimi, B., Schneider, P., & Eifler, T. 2008, A&A, 477, 43
Joachimi, B., Mandelbaum, R., Abdalla, F. B., & Bridle, S. L. 2011, A&A, 527, A26

Joudaki, S., Blake, C., Johnson, A., et al. 2018, MNRAS, 474, 4894
Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, A&A, 638, L1
Kaiser, N. 1992, ApJ, 388, 272
Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92
Kaufman, G. M. 1967, Some Bayesian Moment Formulae, Report No. 6710, Center for Operations Research and Econometrics, Catholic University of Louvain, Heverlee, Belgium
Kayo, I., Taruya, A., & Suto, Y. 2001, ApJ, 561, 22
Kazin, E. A., Sánchez, A. G., & Blanton, M. R. 2012, MNRAS, 419, 3223
Kilbinger, M., & Schneider, P. 2004, A&A, 413, 465
Kitaura, F.-S., Rodríguez-Torres, S., Chuang, C.-H., et al. 2016, MNRAS, 456, 4156
Köhlinger, F., Viola, M., Joachimi, B., et al. 2017, MNRAS, 471, 4412
Köhlinger, F., Joachimi, B., Asgari, M., et al. 2019, MNRAS, 484, 3126
Kohonen, T. 1982, Biol. Cybern., 43, 59
Kosowsky, A., Milosavljevic, M., & Jimenez, R. 2002, Phys. Rev. D, 66, 063007
Krause, E., & Eifler, T. 2017, MNRAS, 470, 2100
Krause, E., & Hirata, C. M. 2010, A&A, 523, A28
Krause, E., Eifler, T. F., Zuntz, J., et al. 2017, ArXiv e-prints [arXiv:1706.09359]
Kuijken, K. 2011, The Messenger, 146, 8
Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2
Kunz, M., Trotta, R., & Parkinson, D. R. 2006, Phys. Rev. D, 74, 023503
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Lemos, P., Köhlinger, F., Handley, W., et al. 2020, MNRAS, 496, 4647
Lewis, A., & Bridle, S. 2002, Phys. Rev. D, 66, 103511
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Li, Y., Hu, W., & Takada, M. 2014, Phys. Rev. D, 89, 083519
Lin, C.-H., Harnois-Déraps, J., Eifler, T., et al. 2020, MNRAS, 499, 2977
Loverde, M., & Afshordi, N. 2008, Phys. Rev. D, 78, 123506
LSST Dark Energy Science Collaboration 2012, ArXiv e-prints [arXiv:1211.0310]
Mandelbaum, R. 2018, ARA&A, 56, 393
Mandelbaum, R., Hirata, C. M., Seljak, U., et al. 2005, MNRAS, 361, 1287
Masters, K. L., Maraston, C., Nichol, R. C., et al. 2011, MNRAS, 418, 1055
McFarland, J. P., Verdoes-Kleijn, G., Sikkema, G., et al. 2013, Exp. Astron., 35, 45
Mead, A. J., Peacock, J. A., Heymans, C., Joudaki, S., & Heavens, A. F. 2015, MNRAS, 454, 1958
Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315
Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Park, Y., Rozo, E., & Krause, E. 2021, Phys. Rev. Lett., 126, 021301
Percival, W. J., Ross, A. J., Sánchez, A. G., et al. 2014, MNRAS, 439, 2531
Planck Collaboration V. 2020, A&A, 641, A5
Planck Collaboration VI. 2020, A&A, 641, A6
Price, M. A., McEwen, J. D., Pratley, L., & Kitching, T. D. 2021, MNRAS, 500, 5436
Raveri, M., & Hu, W. 2019, Phys. Rev. D, 99, 043506
Reid, B., Ho, S., Padmanabhan, N., et al. 2016, MNRAS, 455, 1553
Richard, J., Kneib, J. P., Blake, C., et al. 2019, The Messenger, 175, 50
Robert, C. 2001, The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, Springer Texts in Statistics (New York: Springer)
Ross, A. J., Beutler, F., Chuang, C.-H., et al. 2017, MNRAS, 464, 1168
Rozo, E., Rykoff, E. S., Abate, A., et al. 2016, MNRAS, 461, 1431
Salazar-Albornoz, S., Sánchez, A. G., Grieb, J. N., et al. 2017, MNRAS, 468, 2938
Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2018, MNRAS, 475, 4524
Samuroff, S., Blazek, J., Troxel, M. A., et al. 2019, MNRAS, 489, 5453
Sánchez, A. G., Scoccimarro, R., Crocce, M., et al. 2017, MNRAS, 464, 1640
Schmidt, F., Rozo, E., Dodelson, S., Hui, L., & Sheldon, E. 2009, ApJ, 702, 593
Schneider, P., & Hartlap, J. 2009, A&A, 504, 705
Schneider, P., van Waerbeke, L., Jain, B., & Kruse, G. 1998, MNRAS, 296, 873
Schneider, P., van Waerbeke, L., & Mellier, Y. 2002a, A&A, 389, 729
Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002b, A&A, 396, 1
Schneider, A., Stoira, N., Refregier, A., et al. 2020, JCAP, 2020, 019
Scoccimarro, R. 2004, Phys. Rev. D, 70, 083007
Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, ApJ, 546, 20
Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, ApJ, 859, 101
Sellentin, E., & Heavens, A. F. 2018, MNRAS, 473, 2355
Sellentin, E., Heymans, C., & Harnois-Déraps, J. 2018, MNRAS, 477, 4879
Shanks, T., Metcalfe, N., Chehade, B., et al. 2015, MNRAS, 451, 4238
Sheldon, E. S., & Huff, E. M. 2017, ApJ, 841, 24
Shirasaki, M., & Takada, M. 2018, MNRAS, 478, 4277
Silverman, B. W. 1986, Density Estimation for Statistics and Data Analysis (London: Chapman and Hall)
Simon, P., & Hilbert, S. 2018, A&A, 613, A15
Singh, S., Mandelbaum, R., Seljak, U., Slosar, A., & Vazquez Gonzalez, J. 2017, MNRAS, 471, 3827
Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. 2002, J. R. Stat. Soc. B, 64, 583
Stölzner, B., Joachimi, B., Korn, A., Hildebrandt, H., & Wright, A. H. 2020, A&A, submitted [arXiv:2012.07707]
Takada, M., & Hu, W. 2013, Phys. Rev. D, 87, 123504
Takada, M., & Jain, B. 2004, MNRAS, 348, 897
Takada, M., Ellis, R. S., Chiba, M., et al. 2014, PASJ, 66, R1
Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152
Taruya, A., Bernardeau, F., Nishimichi, T., & Codis, S. 2012, Phys. Rev. D, 86, 103528
Thiele, L., Duncan, C. A. J., & Alonso, D. 2020, MNRAS, 491, 1746
Thomas, D. B., Whittaker, L., Camera, S., & Brown, M. L. 2017, MNRAS, 470, 3131
Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878
Tröster, T., Sánchez, A. G., Asgari, M., et al. 2020a, A&A, 633, L10
Tröster, T., Asgari, M., Blake, C., et al. 2020b, A&A, submitted [arXiv:2010.16416]
Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018a, Phys. Rev. D, 98, 043528
Troxel, M. A., Krause, E., Chang, C., et al. 2018b, MNRAS, 479, 4998
Unruh, S., Schneider, P., Hilbert, S., et al. 2020, A&A, 638, A96
Vale, C., Hoekstra, H., van Waerbeke, L., & White, M. 2004, ApJ, 613, L1
van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, A&A, 642, A200
van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, MNRAS, 476, 4662
Verde, L., Treu, T., & Riess, A. G. 2019, Nat. Astron., 3, 891
White, M., Blanton, M., Bolton, A., et al. 2011, ApJ, 728, 126
Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, A&A, 632, A34
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020a, A&A, 640, L14
Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020b, A&A, 637, A100
Xavier, H. S., Abdalla, F. B., & Joachimi, B. 2016, MNRAS, 459, 3693
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astron. Comput., 12, 45
Zuntz, J., Sheldon, E., Samuroff, S., et al. 2018, MNRAS, 481, 1149

Appendix A: Additional tables related to signal modelling

Table A.1. Choice of fiducial model parameters.

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

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

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

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

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

Appendix B: Magnification bias

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

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

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

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

$$\alpha^{L1}_{mag} = 1.80 \pm 0.15\,; \quad \alpha^{L2}_{mag} = 2.62 \pm 0.28, \tag{B.1}$$

where the errors are the standard deviations of the magnitude counts among the bins selected for the slope measurements.

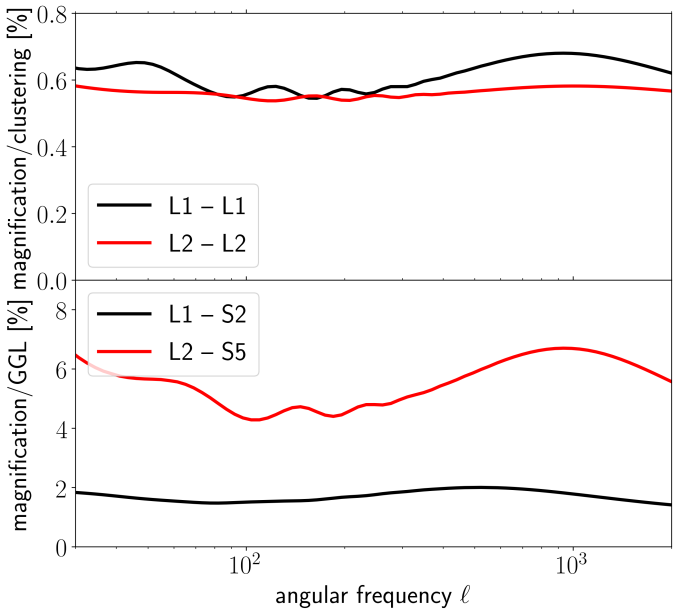

Fig. B.1. Magnification bias contribution relative to galaxy clustering (*top panel*) and the galaxy-galaxy lensing signal (*bottom panel*) for the redshift bin combinations indicated in the legends (cf. Table 2). Shown are angular power spectrum models assuming linear galaxy bias using the values from Table A.1.

Rather than pure statistical noise, these errors therefore have contributions from the limited accuracy of the assumption that the BOSS-selected samples have a simple power-law scaling in the magnitude range that we consider.

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

Appendix C: Propagation of multiplicative shear calibration

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

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

working in the absence of additive biases such as PSF contamination. Here the unsheared ellipticity ϵ is uncorrelated with γ (i.e. there are no intrinsic alignments), and is a combination of the true intrinsic ellipticity $\epsilon^{\rm int}$ and random measurement noise. As noise increases in the imaging, galaxy shapes become increasingly round and our ability to recover the underlying shear decreases, independently of the shear estimation method used. We express this inherent fundamental noise bias in terms of a multiplicative shear calibration correction m that can be seen as both a correction that provides an unbiased shear estimator, and also a weight, reflecting that some galaxies carry little shear information. Regarding this latter point, using the term 'Responsivity' or 'Sensitivity' to describe $(1+m)$ is useful (see e.g., Huff & Mandelbaum 2017; Sheldon & Huff 2017; Zuntz et al. 2018) as this term quantifies how responsive, or sensitive, the observed galaxy is to an induced shear. Unfortunately we cannot measure m precisely for each galaxy and can only determine an estimate $\hat{m}$, where

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

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

We note in passing that our multiplicative shear bias calibration takes the impact of variable survey depth on shear measurement into account because $\hat{m}$ is a function of the signal-to-noise and size of the galaxy image, simulated for a range of different seeings (see Giblin et al. 2021 for details). While we do not take into account that in a poor-seeing, low-depth region the fraction of blended galaxies will be higher than average, Kannawadi et al. (2019) showed the calibration to be robust to changes in galaxy background density and thus the levels of blending well within the derived calibration uncertainty. We therefore do not expect the coupling of blending-induced shear bias with variable depth to make a significant impact on the parameter constraints, as artificially changing the blend fraction across the whole simulated survey changes $\hat{m}$ by an insignificant amount.

C.1. Shear correlation function estimator

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

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

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

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

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

where $W_i := w_i(1+m_i)$ and $\widehat{W}_i := w_i(1+\hat{m}_i)$. The first term introduces shape noise into our analysis (which we explore in Appendix C.3). The second term is the true weighted cosmic shear signal that we wish to extract, where we can see that our

weighted source galaxy distribution now includes the Responsivity, correctly down-weighting galaxies that are unresponsive to shear (as these typically have negative values for m).

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

$$K(\theta) := \sum_{i\in I, j\in J} \widehat{W}_i \widehat{W}_j \, \Delta_{ij}(\theta) \approx (1 + \langle m\rangle_I)\,(1 + \langle m\rangle_J) \sum_{i\in I, j\in J} w_i \, w_j \, \Delta_{ij}(\theta), \tag{C.5}$$

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

C.2. Redshift distribution

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

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

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

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

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

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

C.3. Shape noise estimates

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

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

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

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

where we have approximated the intrinsic ellipticity dispersion by the observed ellipticity dispersion as $|\gamma| \ll 1$, that is $W\epsilon \approx w\epsilon^{\rm obs}$. In evaluating $K(0)$ we use the limit $\Delta_{ij}(0) = \delta_{ij}$. The variances of the ellipticities are calculated after the additive bias correction has been applied (as discussed in Giblin et al. 2021).

C.4. Effective galaxy pair count and number density

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

$$\begin{aligned} {\rm Cov}_{\rm G,sn}\left[\xi_+(\theta);\, \xi_+(\theta)\right] &= {\rm Cov}_{\rm G,sn}\left[\xi_-(\theta);\, \xi_-(\theta)\right] \\ &= K^{-2}(\theta) \sum_{ijkl} W_i W_j W_k W_l \, \Delta_{ij}(\theta)\, \Delta_{kl}(\theta) \\ &\quad \times \left\langle \epsilon_{it}\epsilon_{jt}\epsilon_{kt}\epsilon_{lt} + \epsilon_{it}\epsilon_{jt}\epsilon_{k\times}\epsilon_{l\times} + \epsilon_{i\times}\epsilon_{j\times}\epsilon_{kt}\epsilon_{lt} + \epsilon_{i\times}\epsilon_{j\times}\epsilon_{k\times}\epsilon_{l\times} \right\rangle \\ &= \frac{\sigma_\epsilon^4}{K^2(\theta)} \sum_{ij} W_i^2 W_j^2 \, \Delta_{ij}(\theta). \end{aligned} \tag{C.10}$$

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

can define an effective number of galaxy pairs in the presence of weights and calibration as

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

such that the calibration correction carries through correctly into the covariance matrix when we measure this effective number of pairs directly from the data. While the first equality in Eq. (C.11) is exact, the second one replaces the unobservable noiseless W with the observable quantity $\widehat{W}$ and is in practice used on the data. The effective number density then follows by setting $\Delta_{ij} \equiv 1$, that is by calculating all galaxy pairs in the survey, and using the relation $N_{\rm pair}^{\rm total} = A_{\rm eff}^2 n_{\rm eff,S}^2$ to obtain:

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

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

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

C.5. Tangential shear case

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

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

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

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

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

Appendix D: Analysis of covariance models

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

D.1. Clustering – weak lensing cross-variance

To demonstrate that the clustering and weak lensing signals are uncorrelated and thus statistically independent (in the Gaussian likelihood approximation), we create 4000 mock realisations with the full BOSS footprint, KiDS, and their overlap. We do not simulate 2dFLenS lens galaxies as the additional GGL measurement outside the area from which clustering was obtained only acts to further reduce correlations. The current FLASK implementation does not allow us to incorporate line-of-sight modes so that we cannot model the redshift-space correlation function. However, since weak lensing only depends on transverse modes of the density distribution, it is sufficient to consider the projected angular correlation function,

$$w^{(ij)}(\theta) = \int_0^\infty \frac{{\rm d}\ell\, \ell}{2\pi}\, {\rm J}_0(\ell\theta)\, C_{\rm gg}^{(ij)}(\ell), \tag{D.1}$$

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

Figure D.1 shows the resulting correlation matrix. Cross-correlations between the clustering correlation functions and any of the weak lensing signals only very rarely exceeds 0.05, and this is largely due to residual noise in the mock covariance. We can therefore safely assume that the clustering and weak lensing parts of the data vector are independent. This trend is driven by the fact that only 3% of the BOSS survey area overlaps with KiDS and thus with the weak lensing measurements. Joint clustering and weak lensing measurements over the same sky area do produce significant cross-correlations (e.g.,

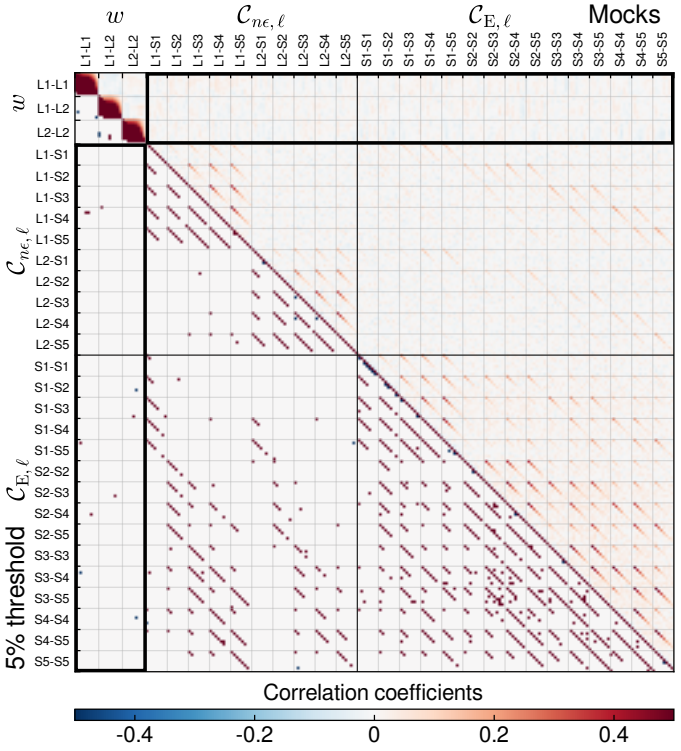

Fig. D.1. Correlation coefficient matrix for the angular galaxy clustering correlation function w, the galaxy-galaxy lensing band power $C_{\rm n\epsilon}$, and the cosmic shear band power C_E calculated over the full BOSS and KiDS-1000 footprints. The upper triangle shows the correlation matrix as calculated from the mocks; the lower triangle indicates correlation coefficients above ±5%. The black rectangles highlight the cross-correlations between galaxy clustering and weak lensing statistics. These are negligible, with few elements fluctuating above the 5% threshold.

Krause & Eifler 2017; Krause et al. 2017) and therefore demand for a more homogeneous approach to summary statistics and their covariance than taken in this work.

D.2. Correlation function covariance

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

Figure D.2 shows the ratio of the square root of the diagonal elements of the mock and analytic covariances for the three survey configurations. Off-diagonals in the form of correlation coefficients are shown in Fig. D.3 for the most realistic Egretta case (the Fourier space analogues of these plots are Figs. 13 and 14). We generally find very good agreement between the mocks and the analytic approach in all cases. Significant deviations are limited to the largest-scale data point in $\xi_\pm$ and to scales larger than 10 arcmin for GGL, with the mock standard deviations up to 20% larger. This under-prediction by the analytic model is because its Gaussian sample variance terms ignore survey boundary effects which enhance them by up to a factor two when switching from a simple rectangular footprint to the realistic KiDS survey geometry (Buceros to Cygnus)[30]. Variable depth effects (Cygnus to Egretta) have negligible impact on the cosmic shear covariance beyond modifications to the galaxy pair counts, but cause a small additional increase in the GGL covariance.

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

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

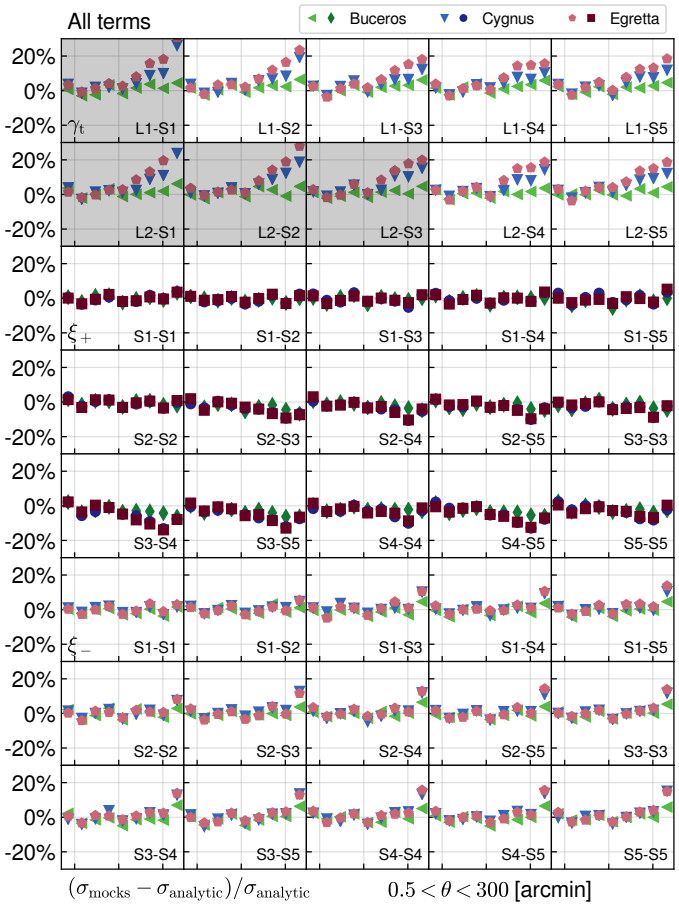

Fig. D.2. Relative difference between the square root of the diagonals of the mock and analytic covariances of the weak lensing correlation functions, i.e. the real-space analogue of Fig. 13. *Top two rows*: GGL signals, *centre three rows*: ξ_+, and *bottom three rows*: ξ_-, with bin combinations indicated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

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

For the interpretation of the signal-only case in Fig. D.5 it is helpful to bear in mind that the only contributing analytic covariance term that is sensitive to survey geometry is the super-sample covariance (SSC). As SSC is suppressed in GGL, the analytic model predicts no geometry dependency, which the mock result suggests is plausible above ca 10 arcmin. Below however, non-Gaussian contributions are seen to cause excess covariance in the

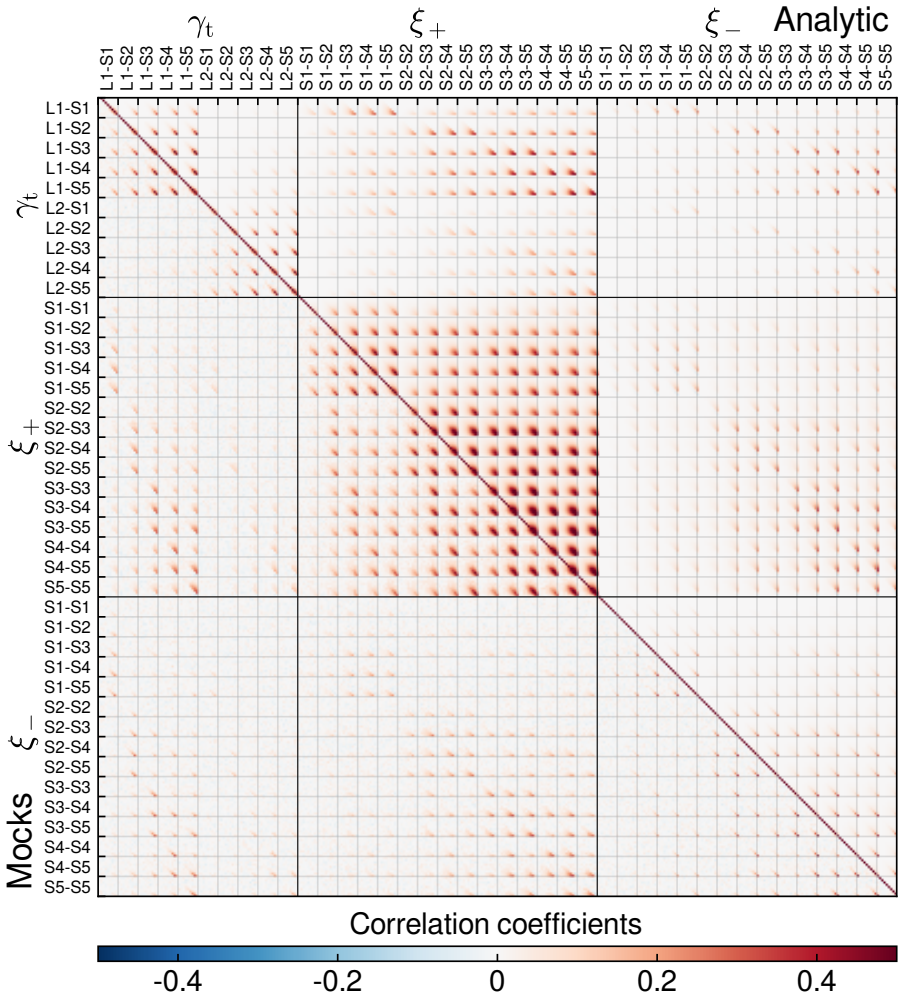

Fig. D.3. Comparison of the correlation coefficients in the weak lensing correlation function covariance between the mocks (lower left) and analytic approach (upper right) in the Egretta setup. As indicated in the labels, large blocks correspond to the correlation functions $\langle\gamma_t\rangle$, ξ_+, and ξ_-, while small blocks correspond to the tomographic bin combinations.

Egretta case, but it remains unclear to what degree this effect is influenced by the lognormal and linear galaxy bias assumptions, as well as the resolution limit, in our simulations. As regards the cosmic shear signals, mocks and analytic model generally agree fairly well in that the signal-only covariance is suppressed in the Egretta relative to the Buceros case[31], while on large scales discrepancies occur, again due to the neglect of survey geometry effects in the Gaussian sample variance. It is interesting to note that even in the signal-only case the differences in the standard deviation between analytic and mock covariances never exceed 30%.

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

D.3. Band power covariance

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

Figure D.9 highlights the differences in correlation coefficients between the mock and analytic covariance in the Egretta case. For the vast majority of coefficients the differences are consistent with scatter due to the finite number of mock realisations.

[31] While having the same total area, the Egretta footprint is spread out over a larger fraction of the sky, primarily due the separation between KiDS-N and KiDS-S. Therefore, the Egretta footprint is able to accommodate larger modes of the large-scale matter density distribution, thereby decreasing super-sample covariance.

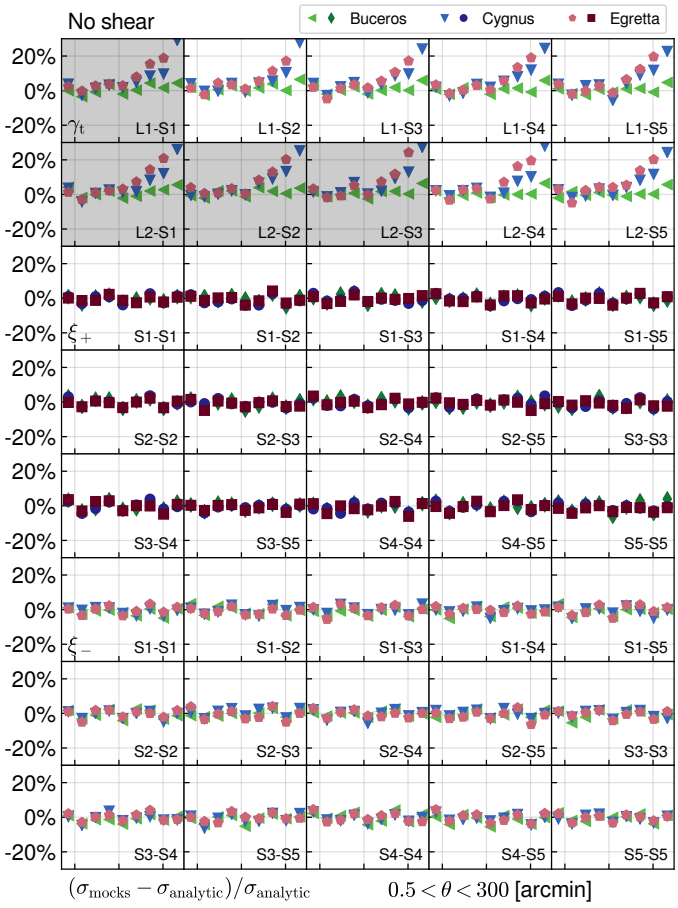

Fig. D.4. Same as Fig. D.2, but with all weak lensing signals removed from both the analytic and mock covariances, i.e. only noise and sample variance due to clustering contribute.

Systematic differences lie in the very thin tails, mostly at the positive end, indicating a larger mock value. These typically occur in the cross-variance between cosmic shear and GGL and are dominated by the mixed term.

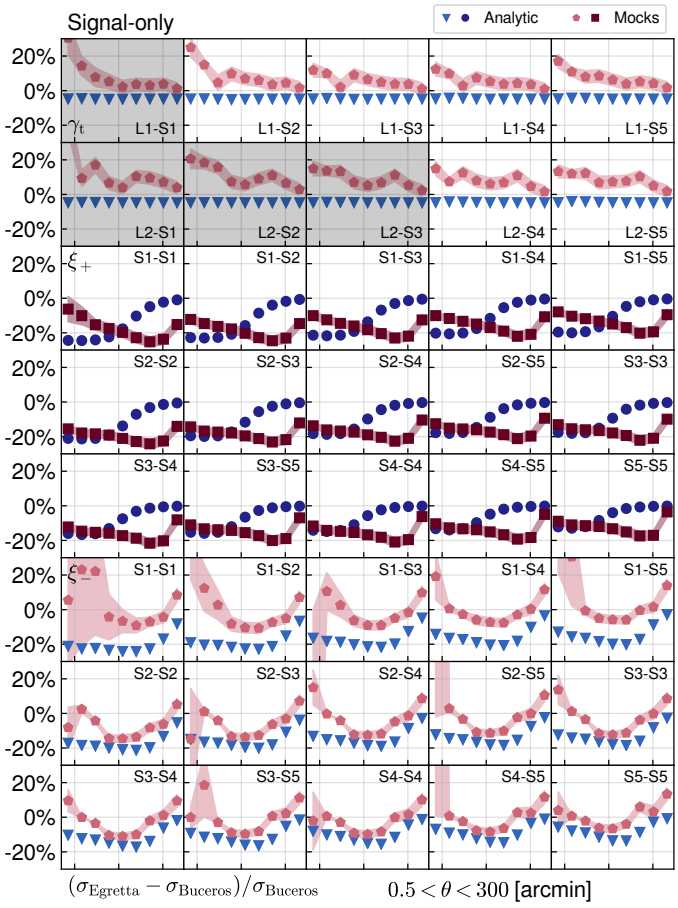

Fig. D.5. Relative difference between the square root of the diagonals in the Egretta (realistic mask and depth variations) and Buceros (rectangular mask, uniform depth) covariances of the weak lensing correlation functions, with all shape noise contributions removed ($\sigma_\epsilon = 0$). Red (blue) symbols show results for the mock (analytic) covariance. Bands around the mock data points indicate the standard error determined from a jackknife estimate of variance. *Top two rows*: GGL signals, *centre three rows*: ξ_+, and *bottom three rows*: ξ_-, with bin combinations indicated in the panels.

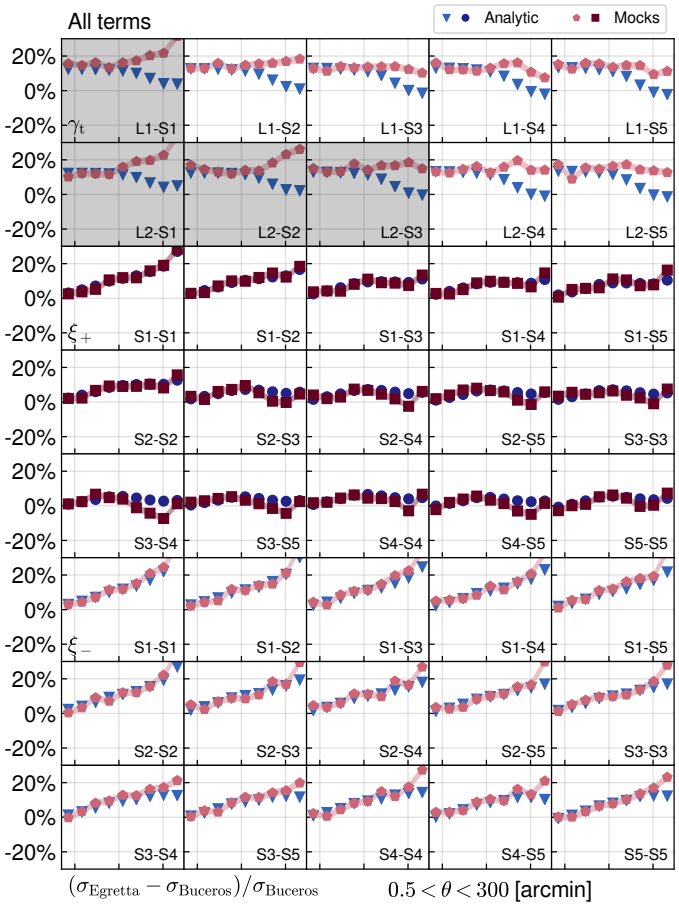

Fig. D.6. Same as Fig. D.5, but for the full covariance with all cosmological contributions included.

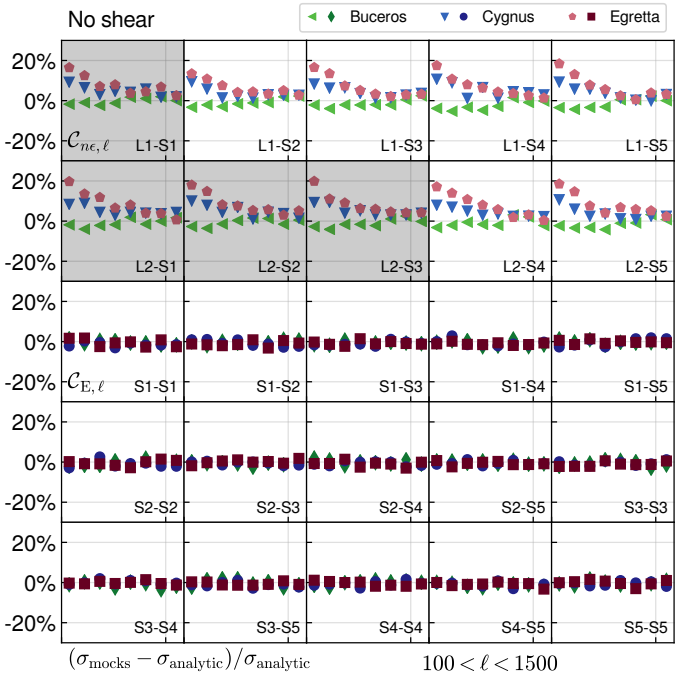

Fig. D.7. Same as Fig. 13, but with all weak lensing signals removed from both the analytic and mock covariances, i.e. only noise and sample variance due to clustering contribute.

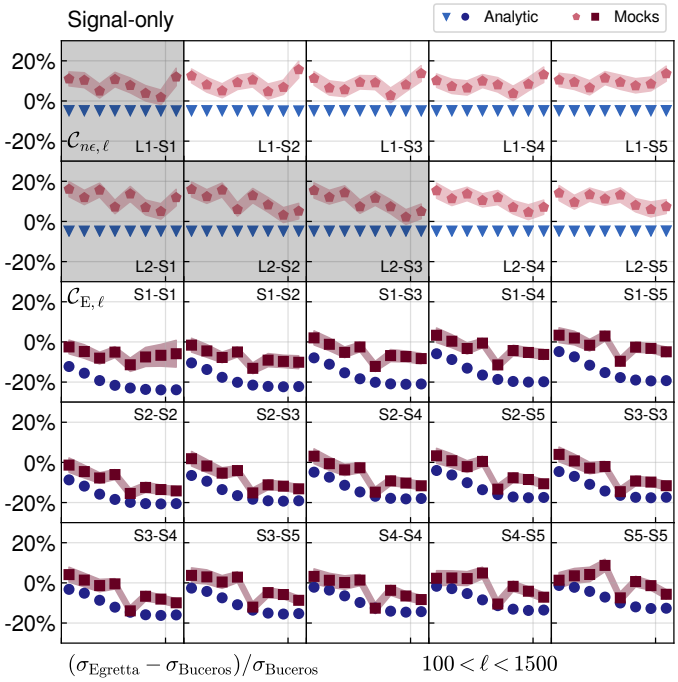

Fig. D.8. Same as Fig. 12, but with all shape noise contributions to the analytic and mock covariances removed ($\sigma_\epsilon = 0$).

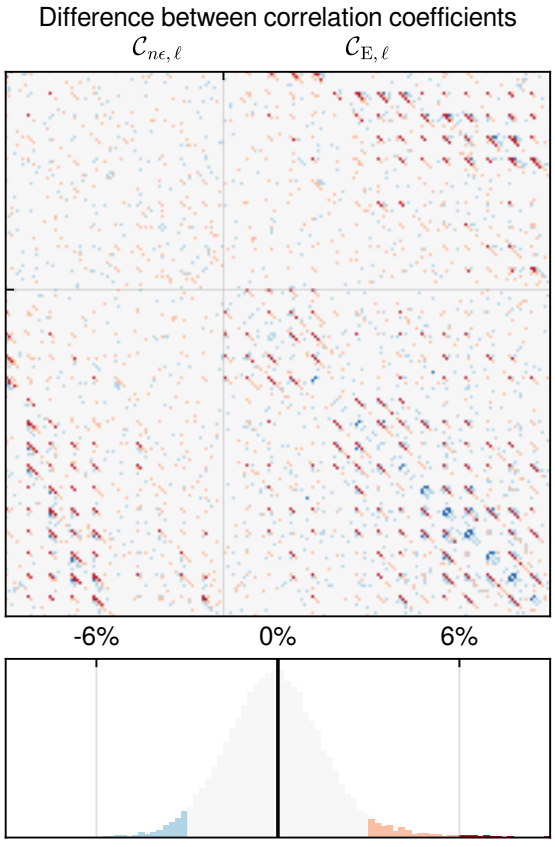

Fig. D.9. Relative difference between the correlation coefficients of the mock and analytic covariances of the weak lensing band power signals. *Bottom*: histogram of the relative differences marking values beyond ±6%(±3%) in dark (light) red and blue. *Top*: band power correlation matrix with elements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

Appendix E: Analytic covariance model

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

E.1. Gaussian real-space covariance

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

$$\mathrm{Cov}_{\mathrm{G,sva}}\left[\Xi_\mu^{(ij)}(\bar\theta_1);\,\Xi_\nu^{(kl)}(\bar\theta_2)\right] = \frac{1}{2\pi A_{\mathrm{max},\mu\nu}}\int_0^\infty \mathrm{d}\ell\,\ell\,\mathcal{K}_\mu(\ell\bar\theta_1)\mathcal{K}_\nu(\ell\bar\theta_2) \times\left\{C^{(ik)}(\ell)C^{(jl)}(\ell)+C^{(il)}(\ell)C^{(jk)}(\ell)\right\}, \tag{E.1}$$

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

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

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

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

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

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

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

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

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

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

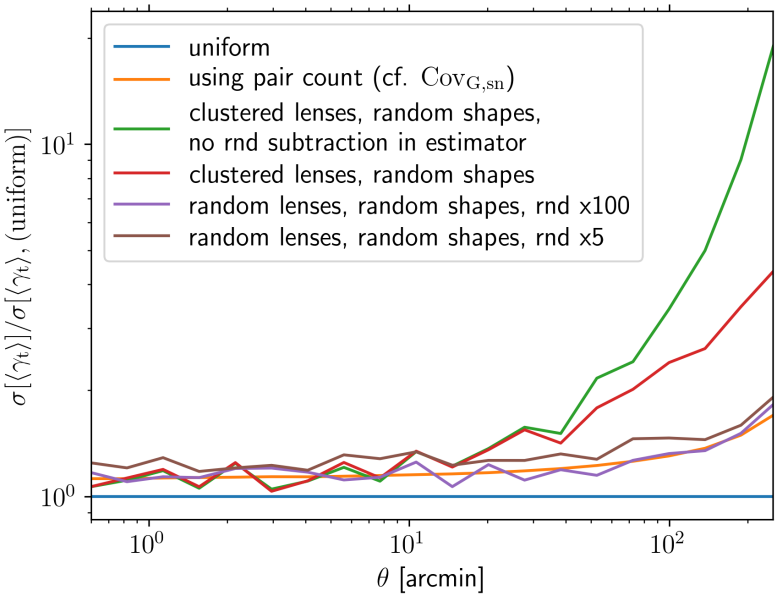

Fig. E.1. Noise contributions to a GGL tangential shear correlation function measurement for BOSS lenses in the KiDS-1000 overlap (L1) and KiDS high-redshift sources (S5). Shown is the standard deviation as a function of angular separation, normalised by the analytic expectation for uniformly distributed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

where we defined

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

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

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

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

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

E.2. Non-Gaussian real-space contributions

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

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

where φ_ℓ is the angle between $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$. The angular trispectrum is calculated by integrating along the line of sight over the matter trispectrum, T_{m}, yielding

$$\begin{aligned} T^{(ijkl)}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4) = \int_0^{\chi_{\mathrm{hor}}} \mathrm{d}\chi\, \frac{W_a^{(i)}(\chi)\, W_b^{(j)}(\chi)\, W_c^{(k)}(\chi)\, W_d^{(l)}(\chi)}{f_{\mathrm{K}}^6(\chi)} \\ \times T_{\mathrm{m}}\left(\frac{\boldsymbol{\ell}_1}{f_{\mathrm{K}}(\chi)}, \frac{\boldsymbol{\ell}_2}{f_{\mathrm{K}}(\chi)}, \frac{\boldsymbol{\ell}_3}{f_{\mathrm{K}}(\chi)}, \frac{\boldsymbol{\ell}_4}{f_{\mathrm{K}}(\chi)}\right) \end{aligned} \tag{E.8}$$

under Limber's approximation. The type of kernel $W_{a,b,c,d}$ is chosen according to the probes Ξ under consideration from among

Eq. (15) for weak lensing and the comoving distance distribution of lens samples plus an expression for galaxy bias for clustering.

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

$$\begin{aligned}\mathrm{Cov}_{\mathrm{SSC}}\left[\Xi_\mu^{(ij)}(\bar\theta_1);\,\Xi_\nu^{(kl)}(\bar\theta_2)\right] = \frac{1}{4\pi^2}\int_0^\infty \mathrm{d}\ell_1\,\ell_1\,\mathcal{K}_\mu(\ell_1\bar\theta_1)\\ \times\int_0^\infty \mathrm{d}\ell_2\,\ell_2\,\mathcal{K}_\nu(\ell_2\bar\theta_2)\\ \times\int_0^{\chi_{\mathrm{hor}}}\mathrm{d}\chi\;\frac{W_a^{(i)}(\chi)\,W_b^{(j)}(\chi)\,W_c^{(k)}(\chi)\,W_d^{(l)}(\chi)}{f_{\mathrm{K}}^6(\chi)}\\ \times\frac{\partial P_{\mathrm{m}}\left[\ell_1/f_{\mathrm{K}}(\chi)\right]}{\partial\delta_b}\;\frac{\partial P_{\mathrm{m}}\left[\ell_2/f_{\mathrm{K}}(\chi)\right]}{\partial\delta_b}\;\sigma^2_{\mathrm{bg},\mu\nu}(\chi),\end{aligned} \tag{E.9}$$

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

$$\sigma^2_{\mathrm{bg},\mu\nu}(\chi) = \frac{1}{A_{\mathrm{eff},\mu}\,A_{\mathrm{eff},\nu}}\sum_\ell P_{\mathrm{m,lin}}\left(\frac{\ell}{f_{\mathrm{K}}(\chi)}\right)\sum_{\mathrm{m}} a^\mu_{\ell m}\,a^{\nu*}_{\ell m}, \tag{E.10}$$

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

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

E.3. Covariance of band powers

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

$$\begin{aligned}\mathrm{Cov}\left[C^{(ij)}_{E/B,m};\,C^{(kl)}_{E/B,n}\right] = \frac{\pi^2}{\mathcal{N}_{\mathrm{m}}\,\mathcal{N}_n}\sum_{a,b}\Delta\theta_a\,\theta_a\,T(\theta_a)\,\Delta\theta_b\,\theta_b\,T(\theta_b)\\ \times\Big\{g^m_+(\theta_a)\,g^n_+(\theta_b)\,\mathrm{Cov}\left[\xi^{(ij)}_+(\theta_a);\,\xi^{(kl)}_+(\theta_b)\right]\\ + g^m_-(\theta_a)\,g^n_-(\theta_b)\,\mathrm{Cov}\left[\xi^{(ij)}_-(\theta_a);\,\xi^{(kl)}_-(\theta_b)\right]\\ \pm g^m_+(\theta_a)\,g^n_-(\theta_b)\,\mathrm{Cov}\left[\xi^{(ij)}_+(\theta_a);\,\xi^{(kl)}_-(\theta_b)\right]\\ \pm g^m_-(\theta_a)\,g^n_+(\theta_b)\,\mathrm{Cov}\left[\xi^{(ij)}_-(\theta_a);\,\xi^{(kl)}_+(\theta_b)\right]\Big\}\end{aligned} \tag{E.11}$$

for cosmic shear, and

$$\begin{aligned}\mathrm{Cov}\left[C^{(ij)}_{\mathrm{n}\epsilon,m};\,C^{(kl)}_{\mathrm{n}\epsilon,n}\right] = \frac{4\pi^2}{\mathcal{N}_{\mathrm{m}}\,\mathcal{N}_n}\sum_{a,b}\Delta\theta_a\,\theta_a\,T(\theta_a)\,\Delta\theta_b\,\theta_b\,T(\theta_b)\\ \times h^m(\theta_a)\,h^n(\theta_b)\,\mathrm{Cov}\left[\langle\gamma_{\mathrm{t}}\rangle^{(ij)}(\theta_a);\,\langle\gamma_{\mathrm{t}}\rangle^{(kl)}(\theta_b)\right],\end{aligned} \tag{E.12}$$

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