KiDS+VIKING-450: improved cosmological parameter constraints from redshift calibration with self-organising maps

Wright, A.H.; Hildebrandt, H.; Busch, J.L. van; Heymans, C.; Joachimi, B.; Kannawadi Jayaraman, A.; Kuijken, K.H.

Astronomy & Astrophysics manuscript no. SOMCosmology
May 11, 2020

KiDS+VIKING-450: Improved cosmological parameter constraints from redshift calibration with self-organising maps

Angus H. Wright[1], Hendrik Hildebrandt[1], Jan Luca van den Busch[1], Catherine Heymans[1,2], Benjamin Joachimi[3], Arun Kannawadi[4], and Konrad Kuijken[5]

[1] Ruhr-Universität Bochum, Astronomisches Institut, German Centre for Cosmological Lensing (GCCL), Universitätsstr. 150, 44801 Bochum, Germany. e-mail: awright@astro.rub.de
[2] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK.
[3] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[4] Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
[5] Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, the Netherlands

Released 12/12/2121

ABSTRACT

We present updated cosmological constraints for the KiDS+VIKING-450 cosmic shear dataset (KV450), estimated using redshift distributions and photometric samples defined using self organising maps (SOMs). Our fiducial analysis finds marginal posterior constraints of $S_8 \equiv \sigma_8 \sqrt{\Omega_\mathrm{m}/0.3} = 0.716^{+0.043}_{-0.038}$; smaller than, but fully consistent with, previous work using this dataset ($|\Delta S_8| = 0.023$). We analyse additional samples and redshift distributions constructed in three ways: excluding certain spectroscopic surveys during redshift calibration, excluding lower-confidence spectroscopic redshifts in redshift calibration, and considering only photometric sources which are jointly calibrated by at least three spectroscopic surveys. In all cases, the method utilised here proves robust: we find a maximal deviation from our fiducial analysis of $|\Delta S_8| \leq 0.009$ for all samples defined and analysed using our SOM. Our largest shift in S_8 is found when calibrating redshift distributions without the DEEP2 spectroscopic subset, where we find $S_8 = 0.707^{+0.046}_{-0.042}$. This difference with respect to the fiducial is both significantly smaller than, and in the opposite direction to, the equivalent shift from previous work. No sample analysed in this work results in a meaningful positive shift in S_8 with respect to our fiducial constraints. These results suggest that our improved cosmological parameter estimates are insensitive to pathological misrepresentation of photometric sources by the spectroscopy used for direct redshift calibration, and therefore that this systematic effect cannot be responsible for the observed difference between S_8 estimates made with KV450 and Planck CMB probes.

Key words. cosmology: observations – gravitational lensing: weak – surveys

1. Introduction

Estimation of cosmological parameters using tomographic cosmic shear requires accurate calibration of source redshift distributions. For Stage III cosmic shear surveys such as the Kilo Degree Survey (KiDS; Kuijken et al. 2019), the Dark Energy Survey (DES; Flaugher et al. 2015), and the Hyper-Suprime Camera Wide-Survey (HSC; Aihara et al. 2018), coherent biases on the order of $\Delta\langle z\rangle = \langle z\rangle_\mathrm{est} - \langle z\rangle_\mathrm{true} \sim 0.04$ are enough to cause significant shifts in estimated cosmological parameter estimates (see, e.g, Hildebrandt et al. 2017). Systematic shifts of this nature are important given the observed (currently mild) tension between cosmological parameters estimated using KiDS weak lensing and cosmic microwave background (CMB) studies (Planck Collaboration et al. 2018). For this reason, considerable effort has been invested in developing, testing, and optimising redshift calibration methodologies for cosmic shear. These methods can typically be grouped into three categories: those which utilise cross-correlation (see, e.g, Schneider et al. 2006; Newman 2008; McQuinn & White 2013; Morrison et al. 2017), stacking of individual redshift probability distributions (see, e.g, Hildebrandt et al. 2012; Hoyle et al. 2018; Tanaka et al. 2018), or direct calibration using spectroscopic redshift training samples (see, e.g, Lima et al. 2008; Hildebrandt et al. 2017, 2020; Buchs et al. 2019; Wright et al. 2020).

The methodological differences, and implicit assumptions, between these estimation/calibration methods mean that they are each susceptible to subtly different biases and systematic effects. For direct calibration methods, the completeness and pre-selection of the spectroscopic training sample has been of particular concern (see, e.g, Gruen & Brimioulle 2017; Hartley et al. 2020). In Wright et al. (2020) we developed an updated implementation of the direct calibration procedure utilising self-organising maps (SOMs; Kohonen 1982), which we found to be less susceptible to bias than previous implementations. We achieved this by the direct flagging and removal of photometric sources which are not directly associated with a spectroscopic calibrator, thereby constructing a sample of fully represented photometric sources and an associated redshift distribution: the 'gold' sample.

In this letter we apply the methodology of Wright et al. (2020) to the KiDS+VIKING-450 dataset of Wright et al. (2019), and perform a tomographic cosmic shear analysis

akin to that of Hildebrandt et al. (2020). The dataset used is described in section 2, as is the definition of our various photometric and spectroscopic analysis samples. Our results are presented in section 3, and we summarise the results presented in this letter in section 4.

2. Dataset and Analysis Methodology

The KiDS+VIKING-450 dataset (hereafter KV450) is presented in Wright et al. (2019), and Hildebrandt et al. (2020, hereafter H20). We utilise the cosmic shear data products from H20 with `lensfit` shape measurements (Miller et al. 2007, 2013), spectroscopic training samples (Vanzella et al. 2008; Lilly et al. 2009; Popesso et al. 2009; Balestra et al. 2010; Le Fèvre et al. 2013; Newman et al. 2013; Kafle et al. 2018), and BPZ photometric redshifts from Benítez (2000), as well as the core of the H20 parameter inference pipeline; we update only the redshift distributions using the new direct redshift calibration methodology of Wright et al. (2020, hereafter W20). Our code is released as a stand-alone analysis package[1], with a wrapper pipeline[2] to perform the analyses presented in this work. We provide the details of these cosmological analysis pipelines in Appendix A.

In this analysis we utilise a range of differently compiled spectroscopic datasets to construct redshift distributions and photometric source 'gold samples' for cosmic shear analysis. A full description of the methods used to construct these redshift distributions and gold samples is presented in W20. Briefly, we utilise self-organising-maps (SOMs), trained on the various spectroscopic datasets, to associate photometric galaxies to spectroscopic galaxies with known redshift. Using these associations, we are able to re-weight the spectroscopic redshift distribution to approximate the (unknown) photometric galaxy redshift distributions. This allows us to flag and remove photometric data which are not associated to spectra (and therefore which are not represented by the re-weighted redshift distributions).

W20 demonstrate that their redshift calibration methodology is less susceptible to systematic biases in redshift distribution reconstruction, when compared with previously incorporated methods used by KiDS (Hildebrandt et al. 2017, H20). Furthermore, using the simulations of van den Busch et al. (in prep.), we are able to estimate biases introduced by calibrating redshift distributions with different spectroscopic calibration samples. We can then use these estimated biases to construct informative priors on the redshift distribution bias parameters (see Appendix B) which we utilise in cosmological parameter estimation.

Finally, the construction of our gold photometric source subsamples requires the simultaneous recalibration of both multiplicative and additive shear measurement bias parameters. While we are able to perform the additive shear bias measurement on-the-fly within our cosmology pipeline, computation of the multiplicative shear biases is more involved. We therefore pre-compute the required multiplicative shear bias values, using the methodology and simulations of Kannawadi et al. (2019), for each of our photometric gold samples. These bias parameters are also given in Appendix B.

[1] https://www.github.com/AngusWright/CosmoPipe
[2] https://www.github.com/AngusWright/CosmoWrapper

2.1. Analysis samples

In this work, we perform cosmic shear parameter estimation using a number of different photometric gold samples, redshift distributions, and priors. Our fiducial analysis defines the gold sample as being those photometric data which are associated with one or more sources within the full KV450 spectroscopic compilation, and whose spectroscopic-to-photometric associations satisfy the quality requirement:

$$|\langle z^s_{\rm spec}\rangle_i - \langle Z^p_{\rm B}\rangle_i| \leq \max\left[5 \times {\rm nMAD}\left(\langle z^s_{\rm spec}\rangle - \langle Z^s_{\rm B}\rangle\right), 0.4\right], \quad (1)$$

for each of the $i \in [1, N]$ association sets, where $z^s_{\rm spec}$ is the spectroscopic redshift of the spectroscopic sources, $Z^s_{\rm B}$ is the photometric redshift of the spectroscopic sources, and $Z^p_{\rm B}$ is the photometric redshift of the photometric sources. This quality requirement filters out associations which have a mean photometric redshift $\langle Z^p_{\rm B}\rangle_i$ (from template fitting with BPZ; Benítez 2000) that catastrophically disagrees with the mean spectroscopic redshift of the association $\langle z^s_{\rm spec}\rangle_i$. This requirement is the same as presented in W20, except that we have imposed a floor on the threshold which defines catastrophic failure; we take as our threshold the maximum of 0.4 and five times the $z^s_{\rm spec} - Z^s_{\rm B}$ dispersion (determined using the normalised median absolute deviation from median; nMAD[3]). Redshift distributions are then calculated per tomographic bin ($Z_{\rm B} \in (0.1, 0.3], (0.3, 0.5], (0.5, 0.7], (0.7, 0.9], (0.9, 1.2]$), as are the photometric gold samples.

In addition to our fiducial analysis, we explore three gold samples constructed from spectroscopic compilations excluding the zCOSMOS, VVDS, and DEEP2 datasets, respectively. We implement these samples both to compare with similar samples run by H20, and to test the sensitivity of our results to pathologically under-representative spectroscopy. Further, we construct one gold sample ('specquality4') using only spectra which have the highest quality flags from their various surveys (referred to as nQ >= 4 spectra, which have $\geq 99.5\%$ confidence), to test the sensitivity of our analysis to spectra with a slightly higher likelihood of catastrophic failures. Finally, we construct a highly restrictive gold sample ('multispec3') which consists only of sources which reside in associations containing spectroscopy from (at least) three different spectroscopic surveys. This selection, coupled with our quality control requirement, essentially restricts our sample to sources whose calibration redshift is supported by multiple spectroscopic surveys with different selection functions, systematic effects, and catastrophic failure modes. This calibration sample is therefore expected to be very robust (albeit at some cost to statistical precision due to a significant reduction in photometric effective number density); mis-calibration of these data would require coordinated catastrophic failure of redshift assignment across multiple spectroscopic campaigns using different instruments and redshifting methods.

3. Results

The results of our various gold sample cosmic shear measurements, quantified using the marginal posterior constraints of the cosmic-shear summary parameter of interest

[3] $\sigma_{\rm nMAD} = 1.4826 \times {\rm med}\,(|x - {\rm med}(x)|)$. The pre-factor ensures normal consistency; that is ${\rm E}[{\rm nMAD}(x_1, .., x_n)] = \sigma$ for $X \sim N(\mu, \sigma^2)$ and large n.

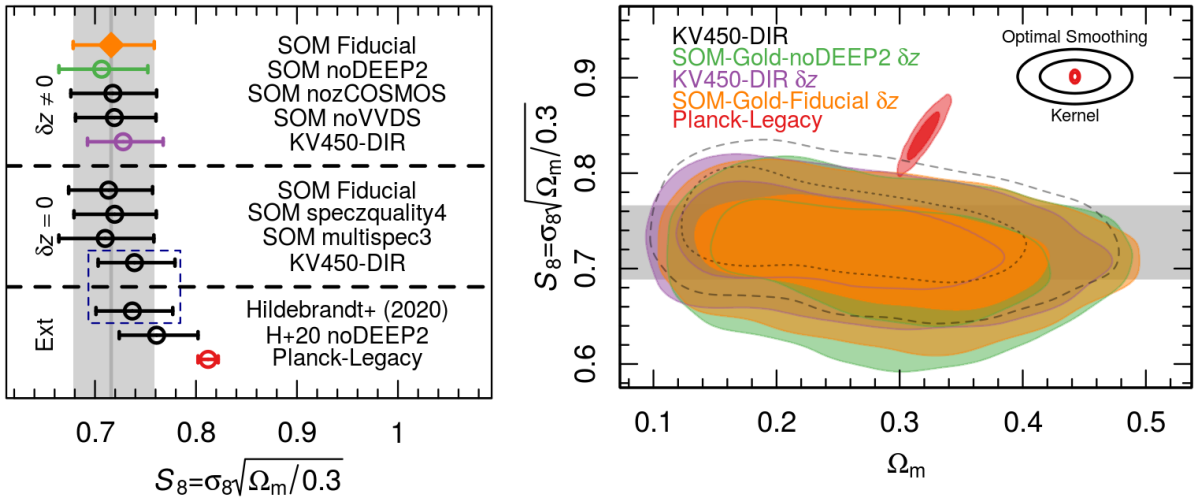

Fig. 1. Posterior constraints of S_8 (left) and Ω_m vs. S_8 (right) for our various gold samples, compared to the results from H20 and Planck CMB. We show results for analyses using updated redshift distribution bias priors ('$\delta z \neq 0$', see Appendix B) and using the fiducial bias priors from H20 ('$\delta z = 0$'). We annotate our contour figure (right) with the two Gaussian smoothing kernels used in generating the contours (one for the cosmic shear contours, and one for the CMB contours). We find that our new cosmology pipeline produces results consistent with the pipeline of H20 (left panel, blue dashed box). Our fiducial results (orange) suggest a slightly lower S_8 than found in previous work: $S_8 = 0.716^{+0.043}_{-0.038}$. When removing various spectroscopic calibration subsamples (DEEP2, VVDS, or zCOSMOS) we find that our constraints of S_8 are extremely stable: $|\Delta S_8| < 0.2\sigma$, demonstrating that the results here are more robust to spectroscopic misrepresentation than previous works. Unlike H20, we find that even pathological misrepresentation at high-redshift ('noDEEP2') is unable to shift our estimates of S_8 to larger values. Performing calibration with only 'certain' redshifts ('speczquality4'; confidence $\geq 99.5\%$) returns S_8 constraints fully consistent with the fiducial, within MCMC noise. Estimating S_8 with photometric sources jointly calibrated by at least three spectroscopic surveys ('multispec3') also gives results fully consistent with the fiducial: $|\Delta S_8| < 0.006$.

$S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$, are shown in Figure 1. Also shown are the results from H20 and Planck-Legacy (Planck Collaboration et al. 2018), for comparison. The left panel is split into 3 sections: analyses performed with non-zero redshift bias priors motivated by the simulations shown in W20 ('$\delta z \neq 0$', see Appendix B), analyses performed with the same (zero-mean) redshift bias priors used by H20 ('$\delta z = 0$'), and external results taken directly from the literature ('Ext'). First, we verify our updated cosmology pipeline by performing an identical cosmological analysis to H20. As seen by the two results highlighted by the blue box in Figure 1, we find that we recover essentially the same S_8 as they report: $S_8 = 0.739^{+0.040}_{-0.037}$ (labelled 'KV450-DIR' in the figure, with $\delta z = 0$) compared to their $S_8 = 0.737^{+0.040}_{-0.036}$ ('Hildebrandt+ (2020)'). We argue that the observed difference ($|\Delta S_8| \lesssim 0.003$) is simply a reflection of noise within our Markov-chain Monte-Carlo (MCMC). For our fiducial gold sample analysis, shown in orange in both panels, we find a marginal constraint of $S_8 = 0.716^{+0.043}_{-0.038}$; smaller than that which was found by H20, but nonetheless fully consistent, especially given that our gold selection produces a different source sample than in H20 which then proceeds through our full analysis pipeline. Additionally, our fiducial analysis is in better agreement with the results of H20 when their dataset and redshift distributions are analysed with modified redshift bias priors (Appendix B): $S_8 = 0.727^{+0.039}_{-0.036}$ ('KV450-DIR' with $\delta z \neq 0$; purple). We observe that our fiducial analysis has a slightly broader marginal S_8 constraint. This is expected when performing our gold selection: by decreasing the size of the photometric dataset which is used for the analysis (which we quantify using the change in the effective number density of cosmic shear source galaxies, $\Delta n_{\rm eff} = n^{\rm gold}_{\rm eff}/n^{\rm all}_{\rm eff} \approx 80\%$ for our fiducial sample; see Appendix C), we increase the statistical noise on our marginal constraints.

We explore the sensitivity of our analysis to the construction of our spectroscopic compilation, by performing our analysis with gold samples constructed without spectra from zCOSMOS, VVDS, and DEEP2. When removing zCOSMOS or VVDS, we find that our marginal constraint on S_8 is unchanged within MCMC noise: $|\Delta S_8| \lesssim 0.003$. In the cases of removing DEEP2 from the calibration sample, we find the greatest shift in our marginal constraint of S_8: $S_8 = 0.707^{+0.046}_{-0.042}$, equating to a shift of $|\Delta S_8| \lesssim 0.2\sigma$. We note though, that (looking at the Ω_m versus S_8 plane) we can see that the shift in S_8 without DEEP2 is driven by an extension of the posterior to lower values, rather than a systematic biasing of the distribution overall.

We draw particular attention to the differences seen between our analysis without DEEP2 and the same analysis performed by H20. When performing their noDEEP2 analysis H20 found a non-trivial increase in S_8 to $S_8 = 0.761^{+0.041}_{-0.037}$ ('H+20 noDEEP2'); a shift of $\Delta S_8 \sim +0.6\sigma$. This difference is attributed, in H20, to a bias in the reconstructed redshift distributions used for this test: removing DEEP2 causes pathological misrepresentation of the high-redshift portion of the spectroscopic colour-colour space, which subsequently causes the reconstructed redshift distributions to be systematically biased low, thereby introducing a positive shift in S_8 for the otherwise unchanged photometric source sample. In the SOM based gold samples we find no such behaviour. We attribute this difference to the gold selection process: our redshift distributions without DEEP2 are similarly skewed low compared to the fiducial (see Appendix C), however unlike H20 our gold selection simultaneously

removes the misrepresented photometric sources. Therefore, while our redshift distributions change significantly between the fiducial and noDEEP2 analyses, both correctly describe the photometric data within their respective gold samples; both are accurate and consistent. We therefore see no biasing of the derived cosmological parameters, but rather just an increase in marginal uncertainties due to the aforementioned decrease in statistical power due to the $\sim 20\%$ reduction in the effective number density of the photometric sample.

In addition to our tests for the effect of pathological colour misrepresentation, we also test the influence of spectra which may have an increased fraction of catastrophic failures. Recall that in our spectroscopic compilation for KV450 we allow only high-confidence ($\geq 95\%$) and/or 'certain' ($\geq 99.5\%$ confidence) spectra; however even high-confidence spectra may have catastrophic failures. In W20 we demonstrated using simulations that expected fractions of catastrophic spectroscopic failures were unlikely to bias calibration of redshift distributions in KV450. Nonetheless, here we explore the influence of the lower-confidence spectroscopy on our conclusions. Our 'speczquality4' gold sample is calibrated using only certain confidence redshifts. The resulting marginal constraints of S_8 differ from our fiducial results only at the level of MCMC noise: $|\Delta S_8| \lesssim 0.003$. We therefore conclude that the presence of lower-confidence spectra in our calibration dataset does not introduce biases in our fiducial marginal constraints of S_8.

Our speczquality4 result is of additional interest in the context of recent work presented by Hartley et al. (2020). For DES (i.e. using fewer photometric bands than used in KiDS), and implementing a redshift calibration methodology akin to that of H20, they find switching between direct calibration using high-confidence ($\geq 95\%$) and certain ($\geq 99.5\%$) spectroscopic samples results in a significant $\Delta\langle z\rangle \geq 0.06$ bias for their highest tomographic bin ($Z_B \in (0.7, 1.3]$). While these biases are not directly applicable to our analysis, any similar systematic bias within our analysis would likely cause a significant change in the estimated cosmological parameters. We find no such systematic bias when switching between direct calibration using high-confidence and certain spectroscopic redshifts, suggesting that this bias is suppressed in our dataset. We hypothesise that this is driven by one, or a combination, of the following three effects. Firstly, that our 9-band photometric space is more resilient to spectroscopic selection biases than the 4-band space considered in Hartley et al. (2020). Secondly, that our deeper and more diverse spectroscopic compilation reduces the sensitivity of the recalibration procedure to strong (survey-specific) spectroscopic selection effects. Finally, that the calibration method of Wright et al. (2020) is more resilient to spectroscopic selection effects than the method used in Hartley et al. (2020). We leave exploration of these three possibilities to future work.

Finally, we extend this test further by implementing more stringent requirements on spectroscopic agreement. Our 'multispec3' gold sample consists only of photometric sources which are calibrated by spectra originating from at least 3 different spectroscopic surveys within our compilation. As stated in Section 2, this requirement places a strong restriction on spectroscopic agreement when coupled with our quality control requirement (Equation 1). For our multispec3 gold sample we find again a result which is consistent with our fiducial analysis: $S_8 = 0.710^{+0.048}_{-0.046}$, corresponding to $|\Delta S_8| \lesssim 0.006$, only slightly larger than the MCMC noise threshold. This slightly stronger deviation is unsurprising, as the multispec3 and noDEEP2 selections remove many of the same photometric sources, as DEEP2 has little redundancy in the spectroscopic compilation (see W20). This result provides a strong indication that the marginal constraints on S_8 presented here are not biased by systematic effects nor catastrophic failures within the spectroscopic calibration sample.

While we have focussed our discussion here on the marginal S_8 constraints, in Appendix D we provide additional marginal constraints for a subset of our posterior parameter distributions and explore other conclusions which we can draw from our gold cosmological analyses, specifically around intrinsic alignments and the posterior probability distributions of Ω_m and σ_8. Briefly, our gold sample marginal show a reduced preference for low values of Ω_m, causing a more consistent recovery of $\Omega_m \approx 0.3$. In all of our gold analyses the marginal constraints are good agreement ($|\Delta X| < 0.2\sigma$ for all parameters X), with the exception of the intrinsic alignment amplitude parameter A_{IA}, which shows up to $|\Delta A_{IA}| \sim 1.0\sigma$ differences among analyses. Importantly, though, our gold sample A_{IA} constraints are all consistent with $A_{IA} = 0$, unlike those from H20, who found $A_{IA} \approx 1$. This updated constraint is in better agreement with recent work on intrinsic alignments (Fortuna et al. 2020), who predict an intrinsic alignment amplitude for KiDS of $0 \leq A_{IA} \leq 0.2$.

4. Summary

We present updated cosmological parameter constraints from the KiDS+VIKING-450 dataset of Wright et al. (2019), estimated using updated redshift distributions following the methodology of Wright et al. (2020). For our fiducial analysis we find a value of S_8 that is smaller than, but nonetheless fully consistent with, the value reported in the previous KiDS+VIKING-450 cosmological analysis of Hildebrandt et al. (2020): $S_8 = 0.716^{+0.043}_{-0.038}$ compared to $S_8 = 0.737^{+0.040}_{-0.036}$ ($|\Delta S_8| \lesssim 0.6\sigma$). We note, however, that when one analyses the data and redshift distributions of Hildebrandt et al. (2020) using updated redshift distribution bias parameters presented in Wright et al. (2020), their S_8 also shifts downward and is in better agreement with our fiducial analysis: $S_8 = 0.727^{+0.039}_{-0.036}$, $|\Delta S_8| \lesssim 0.3\sigma$. We explore the sensitivity of our results to systematic misrepresentation within the spectroscopic calibration dataset by removing multiple spectroscopic subsamples (DEEP2, VVDS, zCOSMOS), each of which uniquely calibrate distinct portions of the colour-redshift space. We find that the results presented here are robust to pathological misrepresentation, whereby even the removal of DEEP2 is unable to cause a significant shift in S_8: $|\Delta S_8| \lesssim 0.2\sigma$. In contrast to work presented by Hartley et al. (2020), we find that our results are unchanged from the fiducial when performing the calibration using only certain (nQ= 4, $\geq 99.5\%$ confidence) spectroscopic redshifts. Finally, we perform an extremely conservative analysis whereby we only consider photometric sources which are simultaneously calibrated by spectra from at least three different spectroscopic surveys; our estimate of S_8 in this case is similarly unchanged: $S_8 = 0.710^{+0.048}_{-0.046}$. Overall these results indicate that, using the redshift calibration methodology of Wright et al. (2020), pathological misrepresentation of photometric sources within the spectroscopic

compilation is not able to produce significant changes in marginal constraints of S_8, and therefore cannot reconcile the $\Delta S_8 \approx 2.5\sigma$ differences observed between cosmological parameters estimated using KiDS and Planck.

Acknowledgements. We acknowledge support from the European Research Council under grant numbers 770935 (AWH, HH, JvdB) and 647112 (CH). HH is also supported by a Heisenberg grant (Hi1495/5-1) of the Deutsche Forschungsgemeinschaft. CH also acknowledges support from the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research. KK acknowledges support from the Humboldt Foundation, and the hospitality of Imperial College London. This work is based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 100.A-0613, 102.A-0047, 179.A-2004, 177.A-3016, 177.A-3017, 177.A-3018, 298.A-5015. The MICE simulations have been developed at the MareNostrum supercomputer (BSC-CNS) thanks to grants AECT-2006-2-0011 through AECT-2015-1-0013. Data products have been stored at the Port d'Informació Científica (PIC), and distributed through the CosmoHub webportal (cosmohub.pic.es).

References

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S. 2013, J. Cosmology Astropart. Phys., 2, 001
Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12
Benítez, N. 2000, ApJ, 536, 571
Buchs, R., Davis, C., Gruen, D., et al. 2019, arXiv e-prints, arXiv:1901.05005
Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
Fortuna, M. C., Hoekstra, H., Joachimi, B., et al. 2020, arXiv e-prints, arXiv:2003.02700
Gruen, D. & Brimioulle, F. 2017, MNRAS, 468, 769
Hartley, W. G., Chang, C., Samani, S., et al. 2020, arXiv e-prints, arXiv:2003.10454
Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, MNRAS, 421, 2355
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338
Kafle, P. R., Robotham, A. S. G., Driver, S. P., et al. 2018, MNRAS, 479, 3746
Kannawadi, A., Hoekstra, H., Miller, L., et al. 2019, A&A, 624, A92
Kohonen, T. 1982, Biological Cybernetics, 43, 59
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
McQuinn, M. & White, M. 2013, MNRAS, 433, 2857
Miller, L., Heymans, C., Kitching, T. D., et al. 2013, MNRAS, 429, 2858
Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315
Morrison, C. B., Hildebrandt, H., Schmidt, S. J., et al. 2017, MNRAS, 467, 3576
Newman, J. A. 2008, ApJ, 684, 88
Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, ArXiv e-prints [arXiv:1807.06209]
Popesso, P., Dickinson, M., Nonino, M., et al. 2009, A&A, 494, 443
Schneider, M., Knox, L., Zhan, H., & Connolly, A. 2006, ApJ, 651, 14
Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9
van den Busch, J., Hildebrandt, H., & Wright, A. H. in prep.
Vanzella, E., Cristiani, S., Dickinson, M., et al. 2008, A&A, 478, 83
Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, A&A, 632, A34
Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020, arXiv e-prints, arXiv:1909.09632

Appendix A: Cosmology and wrapper pipelines

With this letter we release a new implementation of the KiDS cosmological analysis pipeline utilised by H20, which has been generalised for ease of use. The new pipeline, simply called `CosmoPipe`, is available from `https://www.github.com/AngusWright/CosmoPipe`. The package can be installed trivially with the provided master installation script, and generates a clean working environment for each installation of the pipeline to avoid conflicts with, for example, existing python installations.

`CosmoPipe` contains the same analysis steps performed by H20. The pipeline utilises `treecorr` (Jarvis et al. 2004) for computation of cosmic shear correlation functions, and `MontePython` (Audren et al. 2013) for Markov-chain monte-carlo (MCMC) analyses. For clarity, we outline the seven primary steps of the pipeline here.

1. Compute the 2D c-term for all survey patches and tomographic bins;
2. Compute the 1D c-term for all survey patches;
3. Compute 2pt shear correlation functions;
4. Construct the correlation function covariance matrix;
5. Prepare the data for input to `MontePython` MCMC: reformat the correlation functions, reformat the covariance matrix, prepare the montepython likelihood, reformat the Nz distributions, define the correlation function scalecuts, and link any required treecorr files;
6. Run the MCMC;
7. Construct summary figures and statistics from the MCMC chains.

While this pipeline has been largely generalised, it is clear that some of these steps above are tailored for KiDS-like cosmological analyses. For example, the `CosmoPipe` is provided with a version of the public KiDS likelihood that has been pre-formatted to fit seamlessly into the `CosmoPipe`. The code will function equally well with an arbitrary likelihood, albeit with some additional preparation required on the user-side.

Should one wish to perform an analysis such as (or indeed identical to) that presented here, we also provide a wrapper package which links together the cosmological analysis pipeline package and the redshift calibration package of W20. This wrapper package, available at `https://www.github.com/AngusWright/CosmoWrapper`, contains one main script, `Wright2020b.sh`, which performs the entirety of the analysis presented here. This script requires only that the user have the input photometric and spectroscopic calibration datasets supplied, and performs (with one command) the full gambit of analysis required for this letter. These steps include:

1. redshift calibration;
2. gold sample selection;
3. installation of `CosmoPipe`;
4. preparation of `CosmoPipe` for the different gold sample runs;
5. running `CosmoPipe`; and
6. outputting of figures present in this paper.

Some additional input parameters to the `CosmoPipe` are also encoded in the wrapper package, such as the various redshift distribution and multiplicative shear bias priors given discussed in Appendix B.

Appendix B: Gold sample priors

Appendix B.1: Mean redshift distribution biases

We use the method presented in W20 to estimate the redshift calibration biases, using simulations, for each of our gold samples analysed in this work. These bias estimates allow us to create new redshift distribution bias priors for use in our cosmic shear analyses. These priors are presented in Table B.1. As each of the zCOSMOS, VVDS, and DEEP2 samples were simulated by van den Busch et al. (in prep.), we are therefore able to calibrate the redshift bias parameters for our three gold samples which exclude these subsamples: our 'nozCOSMOS', 'noVVDS', and 'noDEEP2' gold samples. Each of these bias estimates allows us to construct informative redshift distribution bias priors per gold sample. To be conservative, we opt to double the uncertainties on the bias found in the simulations when constructing our priors. For samples analysed without modified redshift bias priors (shown throughout this paper as '$\delta z = 0$'), we maintain the redshift bias priors of H20, also shown in Table B.1. We have opted to implement these updated priors for our gold sample analyses, where possible, as they represent our current best-estimate of the true redshift bias parameters inherent to the recalibration method and samples used here, despite the limitations of the simulations used (van den Busch et al. in prep.; Wright et al. 2020). We note, however, that the biases are typically small, being of order $\delta z \lesssim 0.01$ for the majority of samples and bins. Furthermore, in the case of the fiducial analysis, we find that the updated redshift distribution priors make no difference to our conclusions.

Appendix B.2: Multiplicative shear bias

As each of our gold selections produces a different subset of the full photometric sample, this requires a new computation of the multiplicative and additive shear biases, shear correlation functions, covariances, etc. Each of these is incorporated into the pipeline processing, with the exception of the multiplicative shear bias estimation. In H20, multiplicative shear biases are computed using the methodology and simulations of Kannawadi et al. (2019). We invoke the same procedure, thereby generating a bespoke set of multiplicative shear-bias parameters for each of our gold selections, albeit outside of our wrapper pipeline. These bias parameters are given in Table B.2 for each of our gold samples. We note that in all cases we have chosen to implement the same m-bias uncertainty as used in H20: $\Delta m = 0.02$ for all tomographic bins.

We recognise that this implementation of the m-bias estimation may be sub-optimal: the gold selections are strongly colour-dependent, and our current simulation-set for estimating multiplicative bias is entirely monochromatic (using only the r-band imaging and fluxes). However these simulations nonetheless represent the state-of-the-art within KV450, and we leave exploration of how the m-biases change with multi-colour simulations for future studies. These m-bias parameters are required as input to our cosmology pipeline, and so are documented here. Overall, the gold sample multiplicative biases are very similar, with only the multispec3 calibration differing from the fiducial by more than $\Delta m \sim 0.002$. In all cases, the different m-

Table B.1. Updated redshift distribution bias priors parameters used in different gold sample analyses. Priors are Gaussian ($\mu \pm \sigma$). Parameters are determined from the simulations of van den Busch et al. (in prep.) as described in W20, except that we double the simulation bias standard deviations when constructing our priors. For samples where we want to replicate previous analyses, we implement the prior from H20 ('All $\delta z = 0$').

Gold Sample		Tomographic Redshift Bias Prior δz				
		bin1	bin2	bin3	bin4	bin5
		$Z_B \in (0.1, 0.3]$	$(0.3, 0.5]$	$(0.5, 0.7]$	$(0.7, 0.9]$	$(0.9, 1.2]$
$\delta z \neq 0$	Fiducial	0.000±0.010	0.002±0.012	0.013±0.012	0.011±0.008	−0.006±0.010
	KV450-DIR	0.047±0.010	0.025±0.008	0.032±0.010	−0.004±0.008	−0.013±0.008
	NoDEEP2	−0.001±0.010	0.002±0.012	−0.002±0.012	−0.009±0.010	−0.015±0.010
	noVVDS	0.001±0.010	0.001±0.012	0.024±0.014	0.014±0.010	−0.007±0.012
	nozCOSMOS	0.005±0.026	0.005±0.016	0.032±0.014	0.030±0.010	0.002±0.012
All $\delta z = 0$		0.000±0.039	0.000±0.023	0.000±0.026	0.000±0.012	0.000±0.011

bias values are well within the assumed multiplicative bias uncertainty used here.

Appendix C: Gold sample representation statistics

In this work we have tested the sensitivity of our cosmological parameter estimates to differently constructed gold samples within KV450. Each of these gold samples produces a subset of the available photometric data, and results in a different set of tomographic redshift distributions. We present these representation statistics and corresponding redshift distributions means here in Table C.1.

The combinations of mean redshift and representation statistics tells an interesting story regarding which photometric data are being removed by each of our gold sample definitions. There is a clear correlation between the removal of photometric data and a subsequent decrease in the mean redshift of the tomographic bins. The most obvious examples of this are in the cases of our noDEEP2 and multispec3 samples, where the gold selection removes 30% and 45% of the fiducial $n_{\rm eff}$ in the fifth tomographic bin, respectively. These samples also show the largest redshift distribution shifts within our gold samples: $\Delta\langle z\rangle \sim 0.05$ in the fifth tomographic bin. The suggests that the gold-sample definition is preferentially removing truly high-redshift sources from the photometric sample, as expected. This is indicative of the robustness of the joint redshift distribution estimation and gold selection; unlike the case of the redshift calibration in H20, each combination of gold-sample and redshift distribution presented here is compatible, and differences in sample mean redshifts are not indications of bias in the redshift calibration methodology. This is an important distinction between the different redshift distributions presented here and in H20.

Finally, we note the impact that the reduced effective number density of each gold sample has on our posterior constraint of S_8. The multispec3 subsample, for example, has roughly 50% of the photometric $n_{\rm eff}$ of the KV450-DIR sample per tomographic bin, but shows only a $\sim 35\%$ larger uncertainty on S_8. This is in agreement with the results of H20, who found that the KV450-DIR S_8 uncertainty was limited equally by statistical and systematic uncertainties.

Appendix D: Additional marginal constraints

Here we present a subset of the additional posterior marginal constraints from a subset of our gold sample analyses. In Figure D.1 we show four of the 14 cosmological and nuisance parameters which are used by our likelihood model ($A_{\rm IA}$, $n_{\rm s}$, h, and $\ln 10^{10} A_{\rm s}$), as well as three derived parameters ($\Omega_{\rm m}$, σ_8, and S_8). The mean and standard deviations of these posterior distributions are also provided in Table D.1. For an in depth description of the likelihood used here see H20. We have selected these parameters to show as they are of cosmological interest and/or are not prior dominated in our analysis (unlike, e.g., the redshift distribution bias parameters).

The marginal distributions from each of our gold samples in Figure D.1 are in good agreement. Comparing the various gold sample analyses to our two 'KV450-DIR' runs, which use data vectors and redshift distributions equivalent to those in H20, we see some interesting differences. Firstly, we note that the gold samples no longer demonstrate a preference for small values of the matter density parameter, $\Omega_{\rm m} \sim 0.18$. Instead, our gold marginal distributions all peak at values $\Omega_{\rm m} \sim 0.3$, in much better agreement with concordance cosmological parameters. This has a subsequent effect on the marginal distribution of σ_8, causing it to be considerably narrower for our gold analysis than in the KV450-DIR cases; we find $\sigma_8 = 0.762^{+0.070}_{-0.180}$ compared to $\sigma_8 = 0.836^{+0.132}_{-0.218}$.

Finally, looking at the marginal constraints on $A_{\rm IA}$, we see that this parameter shows the greatest variation within our gold sample analyses. Interestingly, though, we note that only the results of 'KV450-DIR' (i.e. KV450-DIR using the fiducial redshift bias priors) demonstrate a preference for non-zero values of $A_{\rm IA}$. In all other cases, the marginal constraints are consistent with $A_{\rm IA} = 0$; in agreement with recent work on intrinsic alignments within KiDS (Fortuna et al. 2020), who advocate $0 \leq A_{\rm IA} \leq 0.2$.

Table B.2. Multiplicative shear bias parameters used for each of our gold sample analyses.

Gold Sample	multiplicative shear bias parameter				
	bin1	bin2	bin3	bin4	bin5
Fiducial	-0.0145 ± 0.0200	-0.0176 ± 0.0200	-0.0125 ± 0.0200	0.0045 ± 0.0200	0.0122 ± 0.0200
NoDEEP2	-0.0137 ± 0.0200	-0.0162 ± 0.0200	-0.0112 ± 0.0200	0.0054 ± 0.0200	0.0130 ± 0.0200
noVVDS	-0.0143 ± 0.0200	-0.0172 ± 0.0200	-0.0116 ± 0.0200	0.0047 ± 0.0200	0.0125 ± 0.0200
nozCOSMOS	-0.0143 ± 0.0200	-0.0159 ± 0.0200	-0.0106 ± 0.0200	0.0053 ± 0.0200	0.0135 ± 0.0200
speczquality4	-0.0141 ± 0.0200	-0.0163 ± 0.0200	-0.0121 ± 0.0200	0.0043 ± 0.0200	0.0125 ± 0.0200
multispec3	-0.0158 ± 0.0200	-0.0203 ± 0.0200	-0.0173 ± 0.0200	-0.0033 ± 0.0200	-0.0012 ± 0.0200

Table C.1. Mean tomographic redshifts and representation statistics of photometric source galaxies within each of our gold samples. Representation is defined using the effective number density of sources for cosmic shear studies, $n_{\rm eff}$, in each of the gold samples relative to a reference sample $n_{\rm eff}$. For the fiducial representation statistic we use the full KV450 photometric dataset for reference (i.e. $\mathrm{n}_{\rm eff}^{\rm fid}/\mathrm{n}_{\rm eff}^{\rm all}$), while all other gold sample representations use the fiducial for reference (i.e. $\mathrm{n}_{\rm eff}^{\rm gold}/\mathrm{n}_{\rm eff}^{\rm fid}$). The statistics are all given per tomographic bin. The table demonstrates that each of our nozCOSMOS, noVVDS, and noDEEP2 gold samples has preferentially removed a different section of the colour-space. This is joined, however, by a shift in the mean redshift of the tomographic bin, indicating that the loss of the colour redshift space has been accounted for in the reconstruction. As expected, the multispec3 selection is highly restrictive, removing $30-45\%$ of the fiducial photometric $n_{\rm eff}$ in every bin.

Gold Sample	$\mathrm{n}_{\rm eff}^{\rm gold}/\mathrm{n}_{\rm eff}^{\rm ref}(\%)$					$\langle z \rangle$				
	bin1	bin2	bin3	bin4	bin5	bin1	bin2	bin3	bin4	bin5
KV450-DIR	100.0	100.0	100.0	100.0	100.0	0.369	0.463	0.643	0.806	0.973
Fiducial	78.6	82.1	79.2	82.3	91.6	0.236	0.379	0.537	0.766	0.948
nozCOSMOS	93.4	92.2	92.0	88.3	91.5	0.214	0.371	0.529	0.755	0.945
noDEEP2	97.7	96.2	88.6	79.5	72.7	0.237	0.374	0.516	0.737	0.908
noVVDS	97.1	92.4	86.2	88.1	91.4	0.237	0.373	0.537	0.766	0.951
speczquality4	95.0	92.0	87.2	86.8	89.4	0.231	0.367	0.524	0.756	0.941
multispec3	71.2	72.7	65.0	55.4	54.5	0.226	0.369	0.515	0.737	0.906

Table D.1. Marginal parameter means and standard deviations for the subset of parameters shown in Figure D.1.

Parameter	KV450-DIR	Fiducial	nozCOSMOS $\delta z \neq 0$	noVVDS	noDEEP2	KV450-DIR $\delta z = 0$
$A_{\rm IA}$	0.282 ± 0.594	-0.344 ± 0.695	-0.366 ± 0.650	0.198 ± 0.665	-0.627 ± 0.775	0.959 ± 0.671
$n_{\rm s}$	1.044 ± 0.130	1.023 ± 0.133	1.042 ± 0.136	1.020 ± 0.137	1.072 ± 0.130	1.032 ± 0.131
h	0.747 ± 0.049	0.742 ± 0.049	0.743 ± 0.049	0.741 ± 0.050	0.741 ± 0.049	0.748 ± 0.048
$\ln 10^{10} A_{\rm s}$	3.158 ± 0.864	2.816 ± 0.806	2.762 ± 0.795	2.617 ± 0.715	2.653 ± 0.743	3.099 ± 0.882
$\Omega_{\rm m}$	0.249 ± 0.082	0.282 ± 0.085	0.286 ± 0.085	0.305 ± 0.085	0.291 ± 0.081	0.259 ± 0.087
σ_8	0.834 ± 0.156	0.768 ± 0.143	0.765 ± 0.145	0.739 ± 0.132	0.743 ± 0.134	0.833 ± 0.160
S_8	0.728 ± 0.035	0.716 ± 0.038	0.718 ± 0.041	0.719 ± 0.038	0.707 ± 0.042	0.739 ± 0.036

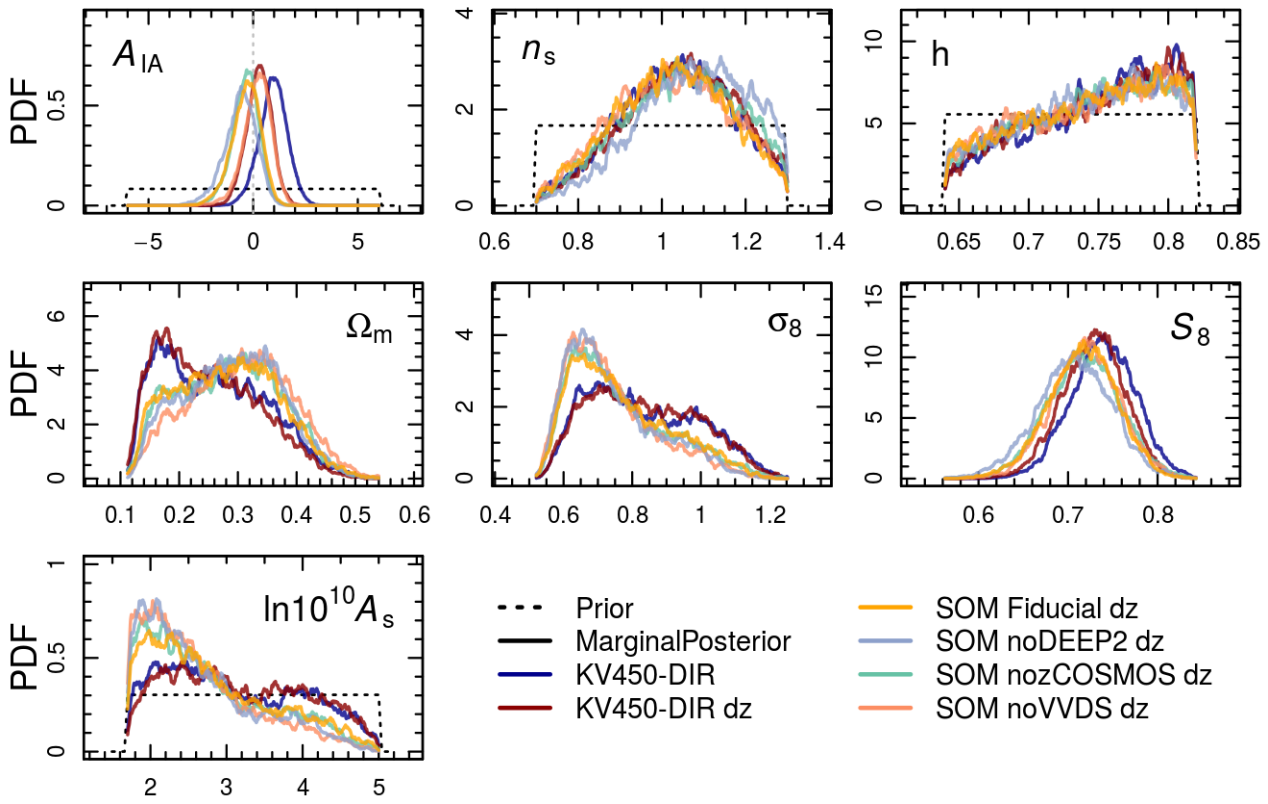

Fig. D.1. Marginal posterior distributions for a subset of the cosmological, nuisance, and derived parameters used in our cosmological model. Coloured lines represent the marginal distributions from various samples. Dashed lines show the priors for all non-derived parameters. There is a clear difference between the marginal distributions of the gold and full-sample ('KV450-DIR') analyses. We note in particular that the previously observed preference within KV450 for small values of the matter density parameter Ω_m is removed in our gold analyses. The gold analyses also prefer a lower value of A_{IA}, consistent with 0 in all cases.