Edinburgh Research Explorer

A Unified Analysis of Four Cosmic Shear Surveys

A Unified Analysis of Four Cosmic Shear Surveys

Chihway Chang,[1,*] Michael Wang,[2] Scott Dodelson,[3] Tim Eifler,[4,5] Catherine Heymans,[6] Michael Jarvis,[7] M. James Jee,[8,9] Shahab Joudaki,[10] Elisabeth Krause,[4,5] Alex Malz,[11,12] Rachel Mandelbaum,[3] Irshad Mohammed,[2] Michael Schneider,[13] Melanie Simet,[4,14] Michael Troxel[15,16] and Joe Zuntz[6]

(The LSST Dark Energy Science Collaboration)

[1]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637*
[2]*Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA*
[3]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[4]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[5]*Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA*
[6]*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*
[7]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[8]*Department of Astronomy, Yonsei University, Yonsei-ro 50, Seoul, Korea*
[9]*Department of Physics, University of California, Davis, California, USA*
[10]*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[11]*Center for Cosmology and Particle Physics, New York University, 726 Broadway, New York, 10003, USA*
[12]*Department of Physics, New York University, 726 Broadway, New York, 10003, USA*
[13]*Lawrence Livermore National Laboratory, Livermore, CA 94551, USA*
[14]*University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA*
[15]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[16]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[*]*e-mail address: chihway@kicp.uchicago.edu*

ABSTRACT

In the past few years, several independent collaborations have presented cosmological constraints from tomographic cosmic shear analyses. These analyses differ in many aspects: the datasets, the shear and photometric redshift estimation algorithms, the theory model assumptions, and the inference pipelines. To assess the robustness of the existing cosmic shear results, we present in this paper a unified analysis of four of the recent cosmic shear surveys: the Deep Lens Survey (DLS), the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS), the Science Verification data from the Dark Energy Survey (DES-SV), and the 450 deg^2 release of the Kilo-Degree Survey (KiDS-450). By using a unified pipeline, we show how the cosmological constraints are sensitive to the various details of the pipeline. We identify several analysis choices that can shift the cosmological constraints by a significant fraction of the uncertainties. For our fiducial analysis choice, considering a Gaussian covariance, conservative scale cuts, assuming no baryonic feedback contamination, identical cosmological parameter priors and intrinsic alignment treatments, we find the constraints (mean, 16% and 84% confidence intervals) on the parameter $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ to be $S_8 = 0.94^{+0.046}_{-0.045}$ (DLS), $0.66^{+0.070}_{-0.071}$ (CFHTLenS), $0.84^{+0.062}_{-0.061}$ (DES-SV) and $0.76^{+0.048}_{-0.049}$ (KiDS-450). From the goodness-of-fit and the Bayesian evidence ratio, we determine that amongst the four surveys, the two more recent surveys, DES-SV and KiDS-450, have acceptable goodness-of-fit and are consistent with each other. The combined constraints are $S_8 = 0.79^{+0.042}_{-0.041}$, which is in good agreement with the first year of DES cosmic shear results and recent CMB constraints from the *Planck* satellite.

1 INTRODUCTION

The large-scale structure of the Universe bends the light rays emitted from distant galaxies according to General Relativity (Einstein

1936). This effect, known as weak (gravitational) lensing, introduces coherent distortions in galaxy shapes, which carry information of the cosmic composition and history.

One of the most common statistics used to extract this infor-

mation is *cosmic shear*, as inferred by the two-point correlation function of galaxy shapes $\xi_{\pm}(\theta)$ (Bartelmann & Schneider 2001). Assuming the flat-sky approximation, these two-point functions are connected to the lensing power spectrum $C(\ell)$ via

$$\xi_{\pm}^{ij}(\theta) = \frac{1}{2\pi} \int d\ell \, \ell J_{0/4}(\theta\ell) C^{ij}(\ell), \qquad (1)$$

where $J_{0/4}$ is the 0th/4th-order Bessel functions of the first kind. The i and j indices specify the two samples of galaxies (or in the case of $i = j$, the galaxy sample) from which the correlation function is calculated. Usually these samples are defined by a certain redshift selection. Under the Limber approximation (Limber 1953; Loverde & Afshordi 2008) and in a spatially flat universe[1], the lensing power spectrum encodes cosmological information through

$$C^{ij}(\ell) = \int_0^{\chi_H} d\chi \frac{q^i(\chi)q^j(\chi)}{\chi^2} P_{NL}\left(\frac{\ell+1/2}{\chi}, \chi\right), \qquad (2)$$

where χ is the radial comoving distance, χ_H is the distance to the horizon, P_{NL} is the nonlinear matter power spectrum, and $q(\chi)$ is the lensing efficiency defined via

$$q^i(\chi) = \frac{3}{2}\Omega_{\rm m}\left(\frac{H_0}{c}\right)^2 \frac{\chi}{a(\chi)} \int_\chi^{\chi_H} d\chi' n^i(\chi') \frac{dz}{d\chi'} \frac{\chi'-\chi}{\chi'}, \qquad (3)$$

where $\Omega_{\rm m}$ is the matter density today, H_0 is the Hubble parameter today, a is the scale factor, and $n^i(\chi)$ is the redshift distribution of the galaxy sample i.

Since the first detection of cosmic shear in Bacon et al. (2000); Kaiser et al. (2000); Wittman et al. (2000); Schneider et al. (2002), the field has seen a rapid growth. In particular, a number of large surveys have delivered cosmic shear results with competitive cosmological constraints in the past few years (Heymans et al. 2013; Becker et al. 2016; Jee et al. 2016; Joudaki et al. 2017a; Troxel et al. 2017; Hildebrandt et al. 2017; DES Collaboration et al. 2017), while ongoing and future surveys will deliver data in much larger volumes and better quality [e.g. the Dark Energy Survey (DES, Flaugher 2005), the Hyper SuprimeCam Survey (HSC, Aihara et al. 2017), the Kilo-Degree Survey (KiDS, de Jong et al. 2015) and the Large Synoptic Survey Telescope (LSST, Ivezić et al. 2008; Abell et al. 2009)].

One of the surprises that has emerged in the past couple of years is that there seems to be a modest level of discordance between different cosmological probes (MacCrann et al. 2015; Freedman 2017; Raveri & Hu 2018). Even though in many of these cases, the level of tension between the different probes still needs to be quantified more rigorously, one consequence has been that the cosmology community has started to more carefully scrutinize how the datasets are analyzed. This is especially important as we expect the statistical power of the datasets to be orders of magnitude better in the near future. If there is indeed a tension between the different probes, it could point to an exciting new direction where the simple ΛCDM cosmology cannot explain all the observables and new physics is needed.

A variety of studies have been carried out to understand systematic effects in weak lensing measurements. This includes systematics from the instrument and the environment, from modeling the point-spread function (PSF) and measuring galaxy shapes, from estimating the redshift of each galaxy, from the theoretical modeling, and many more (see Mandelbaum 2017, and references

therein for a comprehensive list of studies). In this work, we focus on understanding the steps between the shear catalog and cosmological constraints: measuring the shear two-point correlation function [Eq. (1)], estimating the covariance, modeling of the signal, and inferring cosmological parameters. We build a modular and robust pipeline using the PEGASUS workflow engine (Deelman et al. 2015) to analyze the datasets in a streamlined and transparent fashion – this pipeline will serve as the first step towards building up cosmological analysis pipelines for the LSST Dark Energy Survey Collaboration (DESC).

In this paper we apply the pipeline to four publicly available datasets that are precursors to ongoing and future cosmic shear surveys: the Deep Lens Survey (DLS, Jee et al. 2016), the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS, Joudaki et al. 2017a), the Science Verification data from the DES (DES-SV, DES Collaboration et al. 2016), and the 450 deg^2 release of the KiDS (KiDS-450, Hildebrandt et al. 2017). All four surveys were carried out fairly recently and have comparable statistical power, so a uniform pipeline is a powerful way to identify any discrepancies and to understand their origin. A detailed look at the consistency between the four datasets can also inform us about potential systematic issues in the processing that produces the catalogs from which our pipeline begins. It is, however, not the scope of this paper to investigate these issues upstream to our pipeline, where a thorough pixel-level study for each survey may be required.

The paper is organized as follows. In Sec. 2 we describe the details of the four datasets used in this work. In Sec. 3 we describe the pipeline that is used to process the data. We then outline in Sec. 4 the framework in which we compare the datasets and the elements in the pipeline that are allowed to vary. Our results are shown and discussed in Sec. 5 and we conclude in Sec. 6.

2 PRECURSOR SURVEYS

We describe briefly the four datasets used in this work. In Fig. 1 we show the estimated redshift distribution for each dataset. The number of tomographic bins in each case was chosen by the collaboration (and we keep that number fixed throughout), but the range does convey information about the depth of the surveys. For example, the DLS is much deeper and therefore is sensitive to shear at higher redshift. In Fig. 2 we show the footprint of the four datasets on the sky. Since the footprints of these surveys are largely non-overlapping, they can be treated as independent. In Table 1 we list the main parameters used in each of the cosmic shear analyses.

2.1 DLS: the Deep Lens Survey

The DLS (Wittman et al. 2000) consists of five $\sim 2 \times 2$ deg^2 fields that add up to ~ 18 deg^2. Two fields were observed by the Kitt Peak Mayall 4m telescope/Mosaic Prime-Focus Imager (Muller et al. 1998), and the other two by the Cerro Tololo Blanco 4m telescope/Mosaic Prime-Focus Imager. The total DLS dataset was taken over 140 nights of B, V, R and z imaging. The approximate limiting magnitudes for each band (at 5σ) are 26, 26, 27, 26 in B, V, R and z, respectively. The average seeing is $\sim 0.9''$ in R.

The cosmic shear cosmology analysis from DLS was first presented in Jee et al. (2013), and later updated with Jee et al. (2016), which is the analysis we focus on in this paper. The shear measurement method is described in Jee et al. (2013), where an elliptical Gaussian galaxy model is used and image simulations (Jee & Tyson 2011) were employed for calibration of the shear estimate. The

---

[1] For a non-flat universe, one would replace χ by $f_K(\chi)$ in the following equations, where K is the universe's curvature, $f_K(\chi) = K^{-1/2} \sin(K^{1/2}\chi)$ for $K > 0$ and $f_K(\chi) = (-K)^{-1/2} \sinh((-K)^{1/2}\chi)$ for $K < 0$.
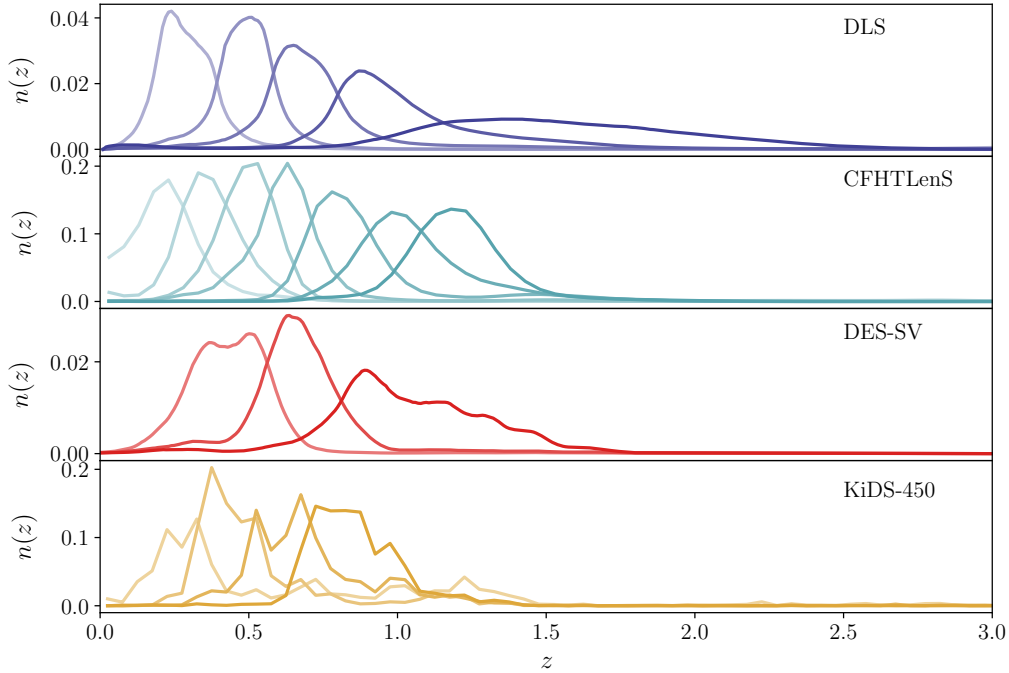
Figure 1. Estimation of the the tomographic redshift distributions used in the four cosmic shear analyses. For DLS, CFHTLenS and DES-SV, stacked photometric redshift probability distribution functions (PDFs) were used; for KiDS-450, the redshift distribution of spectroscopic samples (weighted to match the source galaxies used for the cosmic shear analysis) were used. We see that DLS and CFHTLenS extend to higher redshift compared to DES-SV and KiDS-450.
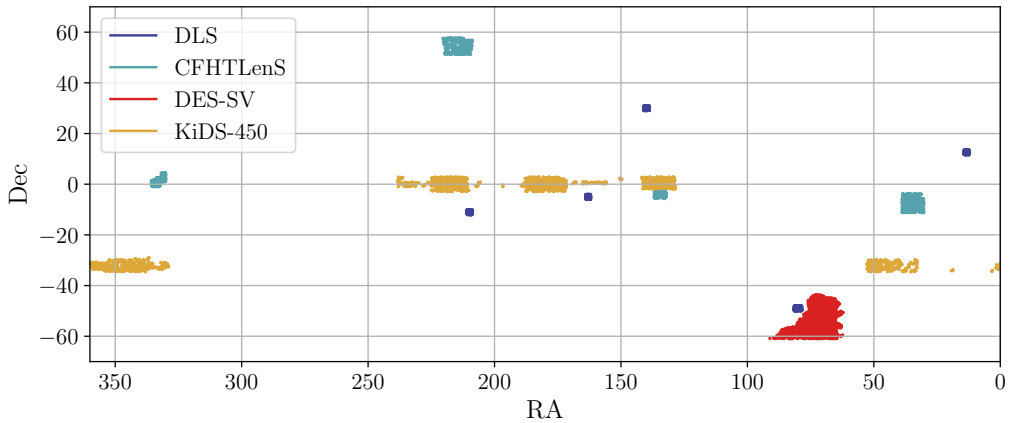
Figure 2. The location and footprints of the four surveys analyzed in this paper. There is essentially no overlap between the footprints of the four surveys, except for a very small part of CFHTLenS and KiDS-450 at (RA, Dec) ≈ (130, -3) deg. We note that the projection in this plot does not reflect the relative area of the four surveys.

photometric redshift (or, photo-z) estimation uses the BPZ code (Benítez 2000) and is validated against the PRIsm MUlti-object Survey (PRIMUS, Coil et al. 2011) in Jee et al. (2013).

DLS is the deepest survey with the smallest area of all the four datasets used in this work. As we do not have access to the shape catalogs for DLS, we start from the pre-measured two-point correlation functions provided by the collaboration.

2.2 CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey

The CFHTLenS data (Erben et al. 2013; Heymans et al. 2012) spans four distinct contiguous fields of approximately 63.8, 22.6, 44.2 and 23.3 deg^2. Images are taken via the Canada-France-Hawaii 3.6m Telescope/MegaCam Imager in six filter bands: u^*, g', r', i', y', z. The limiting magnitudes for each band (at 5σ in $2''$ aperture) are 25.24, 25.58, 24.88, 24.54, 24.71, 23.46 in the six bands, respectively, while the average seeing is $0.68''$ in i', where the shapes are measured.

The cosmic shear cosmology analysis from CFHTLenS was presented first in Fu et al. (2008) and later updated in Heymans

Table 1. Characteristics of the four surveys and some of the modeling choices in each of the analyses.

Parameters	DLS	CFHTLenS	DES-SV	KiDS-450
Reference	Jee et al. (2016)	Joudaki et al. (2017a)	DES Collaboration et al. (2016)	Hildebrandt et al. (2017)
Area ($\deg^2$)	18	94	139	360
Mean redshift	[0.45, 0.62, 0.75, 1.04, 1.51]	[0.35, 0.42, 0.50, 0.64 0.88, 1.06, 1.20]	[0.44, 0.67, 1.03]	[0.50, 0.49, 0.68, 0.85]
σ_e (per component)	[0.25, 0.25, 0.25, 0.25, 0.25]	[0.28, 0.28, 0.28, 0.28, 0.28, 0.28, 0.28]	[0.27, 0.28, 0.28]	[0.29, 0.28, 0.27, 0.28]
n^*_{eff} (arcmin^{-2})	[3.15, 3.58, 3.02, 2.91, 3.75]	[1.28,1.10,1.11,1.59, 2.74,1.75,0.87]	[1.97, 2.03, 2.13]	[2.31,1.83,1.80,1.45]
ξ_+ $\theta_{\min}$ (arcmin)	1.0	1.0	4.6/2.0[†]	0.5
ξ_+ $\theta_{\max}$ (arcmin)	90.0	60.0	60.0	72.0
ξ_- $\theta_{\min}$ (arcmin)	1.0	8.0	56.5/24.5[††]	4.2
ξ_- $\theta_{\max}$ (arcmin)	90.0	120.0	300.0	300.0
Data vector length	240	280	36	130
Covariance	2048 simulations	1988 simulations	126 simulations	analytic
Cosmology inference	CosmoPMC	CosmoMC	CosmoSIS	CosmoMC

[*] We adopt the definition used in Heymans et al. (2012), where $n_{\mathrm{eff}} = A^{-1}(\Sigma_i w_i)^2/\Sigma_i w_i^2$. A is the area of the survey while w_i is the weight for source galaxy i. The summation runs over all source galaxies.

[†] Only for ξ_+^{23} and ξ_+^{33}, the small-scale cutoff is 2 arcmin.

[††] Only for ξ_-^{11} and ξ_-^{12}, the small-scale cutoff is 60 arcmin.

et al. (2013); Kilbinger et al. (2013) and then Joudaki et al. (2017a), which is the focus of this paper. The shear measurement was based on the LensFit package (Miller et al. 2007), which is a likelihood-based model-fitting approach that allows for joint-fitting over multiple observations of the same galaxy. A two-component (disk plus bulge) model is used to fit the galaxy shape and to extract the galaxy ellipticity. The method marginalizes over nuisance parameters such as galaxy position, size, brightness and bulge fraction. Miller et al. (2013) describes the simulation-based calibrations that are applied to the shear catalog. The photo-z estimation was based on the BPZ code (Benítez 2000; Hildebrandt et al. 2012). The catalogs are publicly available[2]. As seen in Fig. 1, the CFHTLenS analysis uses the largest number of tomographic bins.

In addition, in Joudaki et al. (2017a) extensive explorations of the impact of different intrinsic alignment (IA) models, baryonic feedback models, and photo-z uncertainties were performed. When considered independently, only the IA amplitude was found to be substantially favored by the CFHTLenS data. However, with a 2σ negative amplitude, this could be a sign of either simplistic modeling or unaccounted systematics. The CFHTLenS analysis further considered joint accounts of the systematic uncertainties, where the "MIN", "MID" and "MAX" cases included successively conservative treatments of the systematics modeling and scale cuts (along with a "fiducial" case that included no systematics). Joudaki et al. (2017a) found that the S_8 constraints were sensitive to the specific treatment of the systematic uncertainties, where the level of concordance with *Planck* ranged from decisive discordance (MIN) to substantial concordance (MAX). As a result, when quoting the nominal constraints from the collaboration, we show all three cases for CFHTLenS.

2.3 DES-SV: the Dark Energy Survey Science Verification Data

The DES-SV dataset was taken before the official DES run began and was designed to cover a smaller area ($\sim 250\ \deg^2$) to the full depth expected for DES. The area used in the cosmology analysis is a contiguous area of 139 $\deg^2$. Images were taken with the Dark Energy Camera (Flaugher et al. 2015) on the Cerro Tololo Blanco 4m telescope. Five filter bands: g, r, i, z, Y were used to a median depth of $g \sim 24.0$, $r \sim 23.9$, $i \sim 23.0$ and $z \sim 22.3$, respectively. The average seeing is $1.11''$ in r, $1.08''$ in i and $1.03''$ in z – the DES-SV galaxy shapes used information from all three bands.
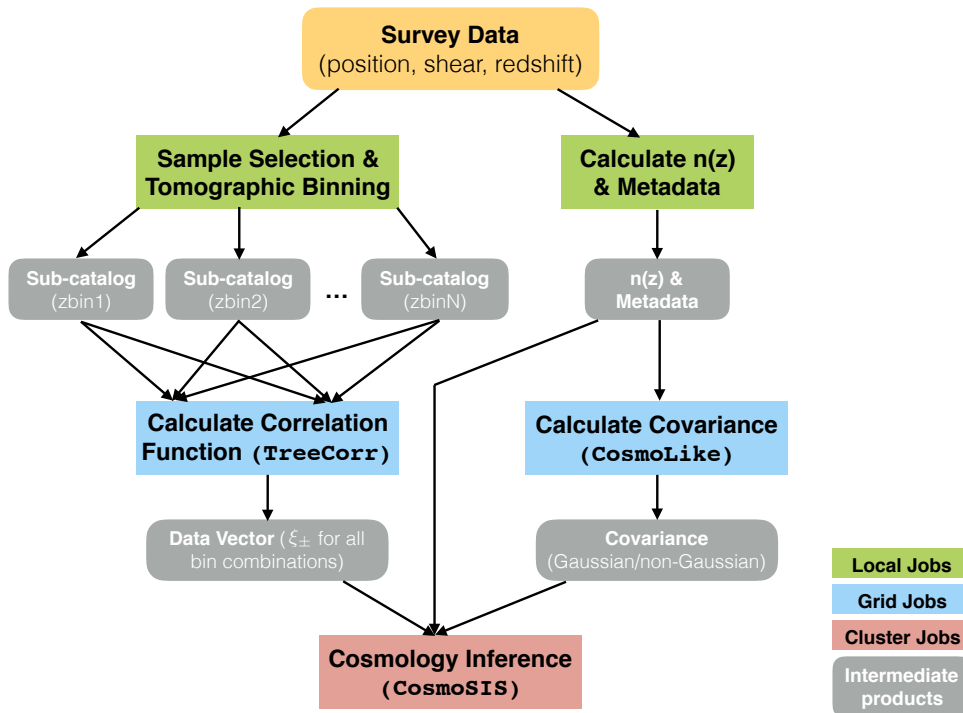
The cosmology analysis from weak lensing was presented in DES Collaboration et al. (2016), while the details and testing of the measurements were recorded in Becker et al. (2016). Two independent shear catalogs were produced from the DES-SV data and have been extensively tested in Jarvis et al. (2016). In this work we use the catalog produced by the shear measurement algorithm NGMIX (Sheldon 2014), which is a fast Bayesian fitting algorithm that models galaxies as a mixture of Gaussian profiles. The Gaussian profiles are chosen to approximate an exponential disk. Several photo-z algorithms were tested in Becker et al. (2016) and Bonnett et al. (2016) including SKYNET (Bonnett 2013) and BPZ (Benítez 2000). In DES Collaboration et al. (2016), results from all shear and photo-z catalogs were presented and shown to be consistent. In this work we use only the NGMIX catalog and the SKYNET photo-z, as these were recommended by DES as the fiducial catalogs with the best performance. All catalogs are publicly available[3].

The analysis pipeline used in DES Collaboration et al. (2016) is based on CoMoSIS (Zuntz et al. 2015), which is the same cosmology inference framework we use in this paper, so we expect very good agreement between our analysis and DES Collaboration et al. (2016).

---

[2] http://www.cfhtlens.org/astronomers/data-store

[3] https://des.ncsa.illinois.edu/releases/sva1/content

Figure 3. Flow chart of steps used in pipeline that goes from survey data to cosmology. The arrow pointing towards the left from the "$n(z)$ & Metadata" box bypasses the covariance calculation – it refers to the route taken when the survey-provided covariances are used in the inference.

2.4 KiDS-450: the 450 deg^2 Kilo-Degree Survey

The KiDS-450 dataset consists of five separate patches covering a total effective area of ~ 360 deg^2. Data was taken using the OmegaCAM CCD Mosaic camera mounted at the Cassegrain focus of the VLT Survey Telescope (VST). There are four SDSS-like filter bands, u, g, r, i, and the image depth is approximately 24.3, 25.1, 24.9, 23.8 in each band, respectively (5σ limit in $2''$ aperture). The median seeing is $0.66''$ in r, and no r-band images have seeing greater than $0.96''$.

The cosmology analysis from cosmic shear using KiDS-450 data was presented in Hildebrandt et al. (2017). The cosmological inference pipeline was largely based on that used in CFHTLenS (Joudaki et al. 2017a), while several updates were made to the measurement pipeline. First, the shear calibration to the LENSFIT shear catalog was based on more sophisticated image simulations (Fenech Conti et al. 2017). Second, a new approach for estimating photo-z and propagating photo-z uncertainties into cosmological inferences was implemented, which we briefly describe below.

The $n(z)$ estimation in KiDS-450 is based on ideas presented in Lima et al. (2008) and implemented in Bonnett et al. (2016). This approach is referred to in Hildebrandt et al. (2017) as the "weighted direct calibration (DIR)" method. The $n(z)$ is taken directly from the redshift distribution of a spectroscopic sample with appropriate re-weighting in the color-magnitude space to correct for the incompleteness and selection effects in both the shear catalog and the spectroscopic sample. Since the $n(z)$'s are derived from a small number of spectroscopic galaxies, they appear more noisy than the other surveys in Fig. 1, where more traditional photo-z methods (stacked redshift probability distribution functions, or PDFs) are used.

3 PIPELINE

A directed acyclic graph (DAG) representing the modular pipeline developed for this analysis is shown in Fig. 3. The pipeline is implemented using the PEGASUS (Deelman et al. 2015) workflow management system. The individual components in the DAG are explained in more detail in Sec. 4, but we outline below the basic structure of the pipeline. Starting from the top, catalogs from each survey are fed into the first two branches of the pipeline which are run in parallel. The first branch (the left half of the DAG) starts with performing sample selection and tomographic binning by sorting catalog data into N_t redshift bins and applying appropriate quality cuts, producing one intermediate catalog file per bin. Next, $N_c = N_t(N_t+1)/2$ jobs are launched in parallel to calculate the two-point shear correlation functions using the TREECORR[4] code. The output of all the parallel jobs are collected to form the data vector for the analysis. The second branch (the right half of the DAG) starts with estimating the full redshift distribution $n(z)$ by summing the redshift PDFs for each individual galaxy[5]. This approach of stacking the redshift PDFs for cosmological inference is not mathematically correct, but is consistent with the implementation of the four surveys under study. The $n(z)$, together with other metadata from each survey (the effective number densities for each tomographic bin, the total shape noise, the survey area) are fed into the calculation of the analytic covariance corresponding to the data vector using the code COSMOLIKE (Krause & Eifler 2016). A to-

---

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

[5] We note that in the later analyses for the individual surveys, some of them do not provide redshift PDFs per galaxy. Instead, they provide the full $n(z)$ for each redshift bin. In those cases (DLS and KiDS) we directly use the survey-provided $n(z)$.

tal of $N_c(2N_c+1)$ COSMOLIKE jobs are launched to calculate each submatrix of the full covariance matrix in parallel. The results for all submatrices are then combined to form the full covariance.

Finally, the outputs from the two branches – the data vector and the covariance matrix – are fed into COSMOSIS (Zuntz et al. 2015) for inference of the cosmological model. The last step also involves choosing the appropriate theory models, priors and scale cuts within COSMOSIS.

This pipeline is written in a modular and generic fashion that strings together the three main codes that are used: TREECORR, COSMOLIKE and COSMOSIS, so that it is easy to substitute different input catalogs, covariances and theory models. Building on this pipeline, it is easy to incorporate other cosmological probes, though that is beyond the scope of this paper. We note also that WLPIPE serves as a test ground for experimenting on different pipeline architecture for future DESC cosmology analyses. For example, we have tested WLPIPE using other workflow engines such as PARSL[6] (Babuji et al. 2018). A similar pipeline was previously constructed for the recent DES Year 1 weak lensing and large-scale structure analyses (DES Collaboration et al. 2017), the cosmic shear part (Troxel et al. 2017) of which was made available to this project. The DES pipeline, however, did not employ any formal workflow management engine. The two pipelines have since been validated against one another to ensure they produce consistent results.

All plots of the cosmological constraints from COSMOSIS chains are plotted using the software package CHAINCONSUMER[7] with setting kde=1.5.

4 COMPARISON FRAMEWORK

The focus of this paper is to compare the cosmic shear analyses of the four precursor surveys in multiple aspects, both within the same dataset and across the four datasets. We describe below the different elements that we consider in this work. We note that as our goal was to investigate and compare the various existing (published) datasets, there was no attempt of blinding throughout the analysis.

4.1 Two-Point Correlation Functions

An important intermediate output of our pipelines is a set of two-point correlation functions for different redshift bins: $\xi_\pm^{ij}(\theta)$ [see Eq. (1)]. These together form the data vector for the cosmological parameter fitting. Except for DLS, whose shape catalogs are not in the public domain yet, we can compare the two-point functions output from WLPIPE with those obtained by the different survey collaborations. For this work, we use the code TREECORR to measure the two-point shear correlation function. TREECORR is a fast tree-based method that allows one to estimate a variety of two- and three-point correlation functions. To measure the two-point shear correlation function, we calculate

$$\xi_\pm^{ij}(\bar{\theta}_\alpha) =$$
$$\frac{\sum_{ab} W_a W_b \left[e_a^t(\vec{\theta}_a) e_b^t(\vec{\theta}_b) \pm e_a^\times(\vec{\theta}_a) e_b^\times(\vec{\theta}_b) \right] \Theta_\alpha(|\vec{\theta}_a - \vec{\theta}_b|)}{\sum_{ab} W_a W_b S_a S_b \Theta_\alpha(|\vec{\theta}_a - \vec{\theta}_b|)}, \quad (4)$$

where e_a^t is the tangential component of the ellipticity of galaxy a with respect to the vector $(\vec{\theta}_a - \vec{\theta}_b)$, and $e_a^\times$ is the cross component; $\bar{\theta}_\alpha$ is the mean angular separation between all galaxy pairs in bin α; W is the weight associated with each galaxy; and S is an (algorithm-dependent) calibration factor defined by each of the different shear catalogs. The last factor, $\Theta_\alpha(|\vec{\theta}_a - \vec{\theta}_b|)$, is 1 when $|\vec{\theta}_a - \vec{\theta}_b|$ is inside angular bin α and 0 elsewhere. When using TREECORR, we set the parameter binslop=0, which means there are no approximations in calculating the angular separation between two galaxies. We note also that DES-SV uses TREECORR to calculate its two-point correlation functions, while for DLS, CFHTLenS and KiDS-450, the measurements are obtained via ATHENA[8].

One final subtle point to note is that $\bar{\theta}$ is the weighted mean of the logarithmic angular separation between all pairs of galaxies in a given angular bin[9], or

$$\bar{\theta}_\alpha = \exp\left[\frac{\sum W_a W_b \ln |\vec{\theta}_a - \vec{\theta}_b| \Theta_\alpha(|\vec{\theta}_a - \vec{\theta}_b|)}{\sum W_a W_b \Theta_\alpha(|\vec{\theta}_a - \vec{\theta}_b|)} \right]. \quad (5)$$

The choice of $\bar{\theta}$ is important because it will be the positions at which the model is evaluated and compared to the $\xi_\pm^{ij}(\theta)$ measurements during the parameter inference process. For DLS, CFHTLenS and KiDS-450, this was not taken into account and the geometric mean of the logarithmic angular bins were used[10]. This will result in a small shift in the parameter inference as we discuss later in Sec. 5.1 and Appendix A. This effect has also been pointed out previously in Joudaki et al. (2018) and Troxel et al. (2018).

4.2 Covariance Matrices

The covariance matrix is an essential element in the pipeline. The full covariance matrix receives contributions from two terms (Cooray & Hu 2001; Sato et al. 2009; Takada & Hu 2013): the Gaussian covariance and the non-Gaussian covariance. The non-Gaussian covariance includes the super-sample covariance (Takada & Hu 2013), which describes the uncertainty induced by large-scale density modes outside the survey window. In this work we use two sets of covariance matrices for each analysis: First, we use the covariance matrices used in the four papers (Jee et al. 2016; DES Collaboration et al. 2016; Joudaki et al. 2017a; Hildebrandt et al. 2017), which were provided by the collaborations. Next, we use a theoretical Gaussian covariance matrix produced by the COSMOLIKE (Krause & Eifler 2016) code. We note that the Gaussian covariance may not be sufficient, especially for DLS, given the smaller area and lower shape noise in this dataset. For further details of the covariance calculation, see Krause et al. (2017). The COSMOLIKE covariance calculation requires the following information from each survey:

- $n(z)$: estimate of redshift distribution for each tomographic bin (see Fig. 1)
- $n_{\rm eff}$: the effective number of source galaxies used in each bin as defined in Heymans et al. (2012) (see Table 1)

---

[6] https://github.com/Parsl/parsl
[7] https://samreay.github.io/ChainConsumer/

[8] http://www.cosmostat.org/software/athena/
[9] The shear response should also be included in the weights, but since the shear response is approximately homogeneous across the survey, we do not incorporate it into this calculation.
[10] For CFHTLenS, $\bar{\theta}_\alpha$ as described in Eq. (5) was used in the early study of Heymans et al. (2013). In the later analysis of Joudaki et al. (2017a), considered in this work, ATHENA had changed this to the geometric mean of the logarithmic angular bins.

- σ_e: standard deviation of the galaxy shape (or, shape noise) for the whole catalog (see Table 1)
- A_{sky}: area of footprint (see Table 1).

Recently, Krause et al. (2017) and Troxel et al. (2018) also pointed out the importance of accounting for the geometry of the footprint, not just its area A_{sky}, by using the survey window function when calculating the analytic covariance. Briefly, one can estimate the effect of the survey geometry by actually counting the number of source galaxy pairs as a function of separation or via an analytic integration of the survey mask. One then uses this information to calculate the shape noise contribution to the covariance instead of the simple geometric calculation based only on the area and mean source number density. We have incorporated this correction to our analytic covariances. Note that this correction does not include the survey geometry correction to the cosmic variance piece of the covariance, which may be important for surveys with low shape noise, such as DLS. The cosmological parameters used to generate all COSMOLIKE Gaussian covariances in this work are: $\Omega_m = 0.286$, $\Omega_b = 0.05$, $\sigma_8 = 0.82$, $h = 0.7$, $n_s = 0.96$.

We first use the survey-provided covariance to check whether we can reproduce the results from the papers. Next we compare the cosmological constraints derived using the survey-provided and the theoretical Gaussian covariance. The four surveys have different approaches to estimate the covariance: for DLS, DES-SV and CFHTLenS, the covariance was estimated via simulations (which are also different between the three cases). For KiDS-450, both simulation and analytic covariances were used and shown to be broadly consistent (though can cause a 1σ shift in the S_8 constraints, see Hildebrandt et al. 2017). The final results were based on the analytic covariance.

For covariances estimated via simulations, we need to apply the Hartlap correction factor H (Hartlap et al. 2007) when inverting the covariance to approximately correct for the bias in the inverse covariance estimate coming from the noise in the simulated covariance C_{sim}. That is,

$$C^{-1} = HC_{sim}^{-1}, \tag{6}$$

where

$$H \equiv \frac{N - p - 2}{N - 1}. \tag{7}$$

N is the number of independent simulations and p is the length of the data vector. We note that this gives an unbiased but still noisy estimate of the inverse covariance. In all our analyses, N is large enough so that the noise associated with the resulting inverse covariance is reasonable (Sellentin & Heavens 2016).

4.3 Cosmological/Nuisance Parameters

In all four survey analyses, of order ten cosmological parameters and parameters modeling systematic effects are varied. The parameters each survey chooses to vary are slightly different and the corresponding priors are also different. In Table 2 we summarize the free cosmological/nuisance parameters and priors for the four analyses under the ΛCDM framework. We note that for CFHTLenS we have chosen to study the "fiducial" setting in Joudaki et al. (2017a), which does not consider any systematic effects. In later analyses when we unify the analysis choices across surveys, the shear calibration bias, photo-z bias and IA amplitude will be allowed to vary. We also note that in Table 2 there are two classes of parametrization of the free cosmological parameters. For DLS and DES-SV,

$[\Omega_m, \Omega_b, h, \sigma_8, n_s]$ was used, whereas for CFHTLenS and KiDS-450, $[\Omega_c h^2, \Omega_b h^2, h, \ln(10^{10}A_s), n_s]$ was used. Here, Ω_b is the baryon density today, h is the unitless Hubble constant ($H_0 = 100h$ km/s/Mpc), σ_8 is the amplitude of the (linear) power spectrum on the scale of 8 h^{-1}Mpc, Ω_c is the cold dark matter density today, A_s is the amplitude of the matter power spectrum, and n_s is the spectral index. Since the priors on the varied parameters are taken to be flat, choosing $\Omega_b h^2$ for example instead of Ω_b translates to choosing a differently shaped prior on the $\Omega_b - h$ parameter space. Furthermore, for CFHTLenS and KiDS-450, the h prior is an indirect one that depends on θ_{MC}, defined as 100 times the ratio of the sound horizon to the angular diameter distance, and is imposed at an intermediate stage. The effective prior on h is therefore not flat. As discussed in Hildebrandt et al. (2017) and Troxel et al. (2018), however, this only has a small effect on the tails of the parameter constraints.

The three classes of nuisance parameters considered here are defined as follows.

- **Intrinsic Alignment:** Most current lensing surveys use the nonlinear alignment model (NLA) proposed by Hirata & Seljak (2004); Bridle & King (2007); Joachimi et al. (2011). The model assumes that the IA power spectra P_{II} and P_{GI} scale with the nonlinear power spectrum P_δ and can be redshift and luminosity-dependent:

$$P_{II}(k,z) = F^2 P_\delta(k,z); \ P_{GI}(k,z) = F P_\delta(k,z), \tag{8}$$

where

$$\begin{aligned} F(z, \bar{L}) &= -A_{IA} C_1 \rho_{crit} \frac{\Omega_m}{D_+(z)} \\ &\times \left(\frac{1+z}{1+z_0} \right)^\eta \left(\frac{\bar{L}}{L_0} \right)^\beta. \end{aligned} \tag{9}$$

Here A_{IA} is a free parameter that dictates the amplitude of the effect, $C_1 = 5 \times 10^{-14} h^{-2} M_\odot^{-1} \text{Mpc}^3$ is a constant, ρ_{crit} is the critical density at redshift zero, and $D_+(z)$ is the linear growth factor that is normalized to 1 today. The power laws η and β determine the redshift and luminosity evolution of the IA effect with z_0 and L_0 chosen as the anchoring redshift and luminosity. $\bar{L}$ is the mean luminosity of the sample. In the four surveys considered in this work, DLS varied A_{IA}, η and β, the "MID" case of CFHTLenS varied A_{IA}, η, while the "MIN" case of CFHTLenS, DES-SV and KiDS-450 only varied A_{IA}[11].

- **Photo-z Uncertainty:** The $n(z)$ estimation can be uncertain and one should marginalize over this uncertainty. We parametrize this uncertainty following the approach used in DLS, CFHTLenS and DES-SV (also see Huterer et al. 2006). That is, we assume the true redshift distribution $n(z)$ has the same shape as the measured redshift distribution $n_{obs}(z)$, but has an uncertain shift in the mean of the distribution, $b_{z,i}$, for each redshift bin i so that

$$n_i(z) = n_{obs,i}(z - b_{z,i}). \tag{10}$$

The approach used in KiDS-450 is slightly different, where the variation in the $n(z)$ itself and the correlation between the errors is accounted for directly. This is done by running a large number (750 is used in Hildebrandt et al. 2017) of chains for each cosmological inference, where each chain uses a different bootstrap sample

---

[11] The "MAX" case of CFHTLenS, DES-SV and KiDS-450 did explore further IA models in DES Collaboration et al. (2016) and Joudaki et al. (2017b), even though they were not taken as the fiducial case.

Table 2. Free parameters in the cosmology inference used in Sec. 5.1, i.e. matching certain cases of the published results as closely as possible. The brackets indicate flat priors with [min, max] and the parentheses indicate Gaussian priors with (mean, standard deviation). We note that for CFHTLenS we choose to use the "fiducial" setting in Joudaki et al. (2017a) as the *Baseline*, which does not consider any systematic effects. In later analyses when we unify the analysis choices across surveys, the shear calibration bias, photo-z bias and IA amplitude will be allowed to vary.

	DLS	CFHTLenS	DES-SV	KiDS-450
Cosmology	$\Omega_{\rm m}$: [0.01, 1.0] $\Omega_{\rm b}$: [0.03, 0.06] σ_8: [0.1, 1.2] h: [0.6, 0.8] n_s: [0.92, 1.02]	$\Omega_{\rm c}h^2$: [0.001, 0.99] $\Omega_{\rm b}h^2$: [0.013, 0.033] $\ln(10^{10}A_s)$: [2.3, 5.0] h: [0.61, 0.81]* n_s: [0.7, 1.3]	$\Omega_{\rm m}$: [0.05, 0.9] $\Omega_{\rm b}$: [0.02, 0.07] σ_8: [0.2, 1.6] h: [0.3, 1] n_s: [0.7, 1.3]	$\Omega_{\rm c}h^2$: [0.01, 0.99] $\Omega_{\rm b}h^2$: [0.019, 0.026] $\ln(10^{10}A_s)$: [1.7, 5.0] h: [0.64, 0.82]* n_s: [0.7, 1.3]
Intrinsic Alignment	A_{IA}: 0.0	A_{IA}: 0.0	A_{IA}: [-5,5]	A_{IA}: [-6,6]
Photo-z bias	b_z: [-0.03, 0.03]†	0	$b_{z,1}$: (0, 0.05) $b_{z,2}$: (0, 0.05) $b_{z,3}$: (0, 0.05)	$b_{z,1}$: (0, 0.036) $b_{z,2}$: (0, 0.015) $b_{z,3}$: (0, 0.01) $b_{z,4}$: (0, 0.006)
Shear calibration bias	m: [-0.03, 0.03]†	0	m_1: (0, 0.05) m_2: (0, 0.05) m_3: (0, 0.05)	m_1: (0, 0.01) m_2: (0, 0.01) m_3: (0, 0.01) m_4: (0, 0.01)

* Priors were placed in an intermediate stage.
† Bins are assumed to be 100% correlated.

of the $n(z)$, and combining all the chains at the very end. As the current WLPipe is not able to implement this operation, we calculate the standard deviation of the mean redshift for each of the 1000 bootstrap $n(z)$'s provided by the collaboration to be [0.036, 0.015, 0.010, 0.006] for each of the redshift bins, and use these values as the priors on the photo-z uncertainty the same way as the other surveys. We find that this approximation gives consistent results to the KiDS-450 approach. The one other subtle point is that in the DLS analysis, the photo-z biases are assumed to be 100% correlated across redshift bins.

• **Shear Calibration Uncertainty:** The shear measurements in each catalog can be uncertain due to imperfect calibration (Mandelbaum et al. 2015). A common way of parametrizing this uncertainty is assuming the true shear γ scales linearly with the measured shear $\gamma_{\rm obs}$ by a factor $(1+m_i)$ for each redshift bin i, plus an additive term c_i (Heymans et al. 2006). That is

$$\gamma_{\rm obs} = \gamma(1+m_i) + c_i. \tag{11}$$

As we will discuss in Sec. 5.1.4, the uncertainty in m_i can either be incorporated at the parameter level or directly in the covariance matrix. We choose the former approach but show that the resulting cosmological constraints are identical (see Fig. B1 in Appendix B). The one other subtle point is that in the DLS analysis, the shear calibration uncertainties are assumed to be 100% correlated across redshift bins. Finally, all surveys we analyzed assume that any residual additive shear biases, c_i, are negligible for the scales used.

In Sec. 5.4, we compare the cosmological constraints from the four datasets using the same priors on cosmological parameters and IA parameters. To see the effect of varying different combinations of the cosmological parameters discussed above, we run analyses for both the DES-SV priors and the KiDS-450 priors. For the photo-z and shear calibration parameters, however, we do not attempt to match between the surveys, as these are parameters that are charac-

terized using the specific datasets. It would be incorrect to assume they have identical priors.

One final subtlety on the modeling side concerns the nonlinear matter power spectrum. Amongst the surveys considered here, DLS uses the Smith et al. (2003) HALOFIT power spectrum, CFHTLenS and KiDS-450 use the HMCODE power spectrum, which is based on Mead et al. (2015), and DES-SV uses the Takahashi et al. (2012) HALOFIT power spectrum. The difference in these power spectrum models can result in slightly shifted cosmological constraints, as discussed in Jee et al. (2016); Joudaki et al. (2017a,b); MacCrann et al. (2015). In this work we use the Takahashi et al. (2012) HALOFIT power spectrum.

4.4 Scale Cuts

In the four cosmic shear analyses, choices were made for which scales will be used for the cosmological inference. The choices were often based on considerations of systematic effects and model uncertainties. In general, the minimum scale is determined by model uncertainties such as baryonic physics and the accuracy of the nonlinear power spectrum. The maximum scale cuts are usually related to survey-specific considerations such as the size of the footprint, additive shear bias, and super-sample covariance. For the four surveys considered, different choices of scale cuts were used and listed in Table 1. A few things to point out: For DLS, the same scale cuts were chosen for ξ_+ and ξ_-, though a discussion of how the scale cuts would change the cosmological constraints was presented in Jee et al. (2016). For CFHTLenS and KiDS-450, in addition to the motivations described above, scales with low signal-to-noise were also removed. Also, the use of smaller scales was justified since the effect of baryonic effects were modeled and marginalized over. For DES-SV, the scale cuts are redshift-dependent and the most conservative.

In our final joint analysis we aim for a uniform scale cut across all four datasets to remove the difference in the four analyses com-

ing from this decision. Since the different surveys have different redshift binning strategies, a unified set of scale cuts is not straightforward. We take the approach of choosing a set of scale cuts in physical units and propagating it into the corresponding angular scale cuts for all of the shear correlation functions. This choice is motivated by the fact that the main consideration that goes into the scale cuts is the uncertainties in the model on small scales (non-linear power spectrum, baryonic effects etc.). The scales on which these effects are important are usually related to the physical size of, for example, dark matter halos. In addition, for cosmic shear measurements, one is not measuring the matter distribution at the redshift of the source galaxies. Instead, it is the matter distribution in the foreground of the source galaxies that we are probing – in specific, matter at the redshift where the lensing efficiency is high [Eq. (3)]. As a result, we choose the scale cuts by calculating the corresponding angular scale cut $\theta_{\min,\pm}$ for some given physical scale $R_{\min,\pm}$ at the redshift of the peak of the lensing kernel z_p. That is, for $\xi_\pm$, we use only angular scales

$$\theta > \theta_{\min,\pm} = \frac{R_{\min,\pm}}{D_A(z_p)}, \tag{12}$$

where $D_A(z_p)$ is the angular diameter distance to redshift z_p.

The physical scale cuts $R_{\min,\pm}$ chosen in our common analysis are $R_{\min,+}$ =1.3 Mpc for ξ_+ and $R_{\min,-}$ =11.4 Mpc for ξ_-. These choices are equal to the most conservative scale cuts amongst the four datasets and very similar to the DES-SV scale cuts. We note that we use $R_{\min,\pm}$ to translate the angular scale cuts between different redshift ranges. The reason for a larger $R_{\min,-}$ is mainly reflecting the difference between the J_0 and J_4 Bessel functions in Eq. (1). We also note that for a more rigorous approach of using truly "linear scale" cuts, see Sec. 3.5 of Joudaki et al. (2017a).

4.5 Cosmological Constraints and Comparisons Metrics

To obtain cosmological constraints, we vary the full set of cosmological and nuisance parameters $\vec{p}$ using a Monte Carlo approach where we assume a Gaussian likelihood, which is the prior multiplied by $e^{-\chi^2/2}$, where

$$\chi^2(\vec{p}) \equiv \sum_{i,j}[d_i - t_i(\vec{p})]\, C^{-1}{}_{ij}\, [d_j - t_j(\vec{p})]. \tag{13}$$

d_i ranges over all data points; $t_i(\vec{p})$ is the theoretical prediction given the set of parameters; and C the covariance matrix. We use the MULTINEST Monte Carlo sampler (Feroz et al. 2009) implemented in COSMOSIS, which has been shown in DES Collaboration et al. (2017) and Krause et al. (2017) to agree very well with other sampling methods such as EMCEE as well as the COSMOLIKE inference code (Krause & Eifler 2017).

This cosmic shear experiments studied in this paper effectively constrains one or at most two cosmological parameters, depending on choices to be discussed below. The parameter that is most tightly constrained is (Jain & Seljak 1997)

$$S_8 \equiv \sigma_8(\Omega_m/0.3)^\alpha, \tag{14}$$

where $\alpha \sim 0.5$ denotes the degeneracy direction in the Ω_m-σ_8 plane so that S_8 gives the most constraining direction of the dataset. The particular value of α depends somewhat on the details of the data and modeling choices. In most existing cosmological analyses, a customary choice is to set $\alpha = 0.5$. However, this could lead to slightly misleading results when comparing different datasets, as not all of them would yield the most constraining S_8 with this

choice of α. In the following analysis, we will use $\alpha = 0.5$ as our fiducial value but discuss in Sec. 5.6 the effect of changing α.

In the next section, we will focus our comparison discussions surrounding four quantities:

- **Signal-to-noise (S/N):** This is simply

$$S/N = \left[\sum_{i,j} d_i C^{-1}{}_{ij} d_j\right]^{0.5}, \tag{15}$$

and it quantifies the statistical significance of the observables.

- **Goodness of fit (χ^2/ν, p.t.e.):** For the best-fit data vector $\hat{D}$, we can calculate the χ^2 per effective number of degree of freedom ν, and the corresponding probability-to-exceed (p.t.e.). It is important to evaluate the goodness of fit for each of the chains in parallel to check for consistency. One disadvantage for using the goodness-of-fit is that the determination of the degree-of-freedom in a high dimensional space is not straightforward. However, for this work the length of the data vector usually dominates over the number of model parameters.

- **1D distance in S_8 (ΔS_8):** We calculate the ratio of the absolute difference between the mean parameter values in the two experiments and the uncertainty in the difference. For two experiments a and b, we thus have

$$\Delta S_8 \equiv \frac{|S_8^a - S_8^b|}{\sqrt{\sigma(S_8^a)^2 + \sigma(S_8^b)^2}}. \tag{16}$$

ΔS_8 can roughly be interpreted as an n-σ difference in S_8 for the two experiments. This metric inherently assumes Gaussianity in the S_8 posterior and ignores possible tensions in other parameter projections. It can also overestimate the inferred disagreement when there are strong degeneracies in other parameter dimensions.

- **Logarithmic Bayes Factor (BF):** Based on Marshall et al. (2006), we consider the logarithmic ratio of the evidence for the two hypotheses: first that the two experiments are measuring the same cosmological parameters and second that they are measuring different cosmological parameters. That is, we calculate

$$\text{BF} = \log_{10}\left(\frac{\int d\Omega P_a(\vec{p})P_b(\vec{p})}{[\int d\vec{p} P_a(\vec{p})][\int d\vec{p} P_b(\vec{p})]}\right). \tag{17}$$

Here the posteriors (including the priors) for each experiment $P_{a,b}$ are integrated over all parameters $\vec{p}$. To properly interpret the BF values, one should evaluate it for cases where the two datasets share the same priors. As a result, we only calculate this at the end of the paper when all analysis choices are unified. We use the criteria BF > -1 to determine whether two surveys are consistent and can be combined. When BF < -1, the Jeffrey scale (Jeffreys 1961) suggests that there is effectively no evidence that the two datasets can be described by the same model.

We note, however, that the BF metric is sensitive to the priors on the constrained parameters, and is usually biased towards consistency (Raveri & Hu 2018).

5 RESULTS

In this section we present the main results of this paper. In Sec. 5.1 we present results from the *Baseline* case: we set out to reproduce the results from the four published papers and discuss in detail the remaining differences between our reproduction and the published results, which we refer to as the *Published Baseline*. We also calculate several comparison metrics in order to understand the internal (external) consistency within (between) the four datasets. In
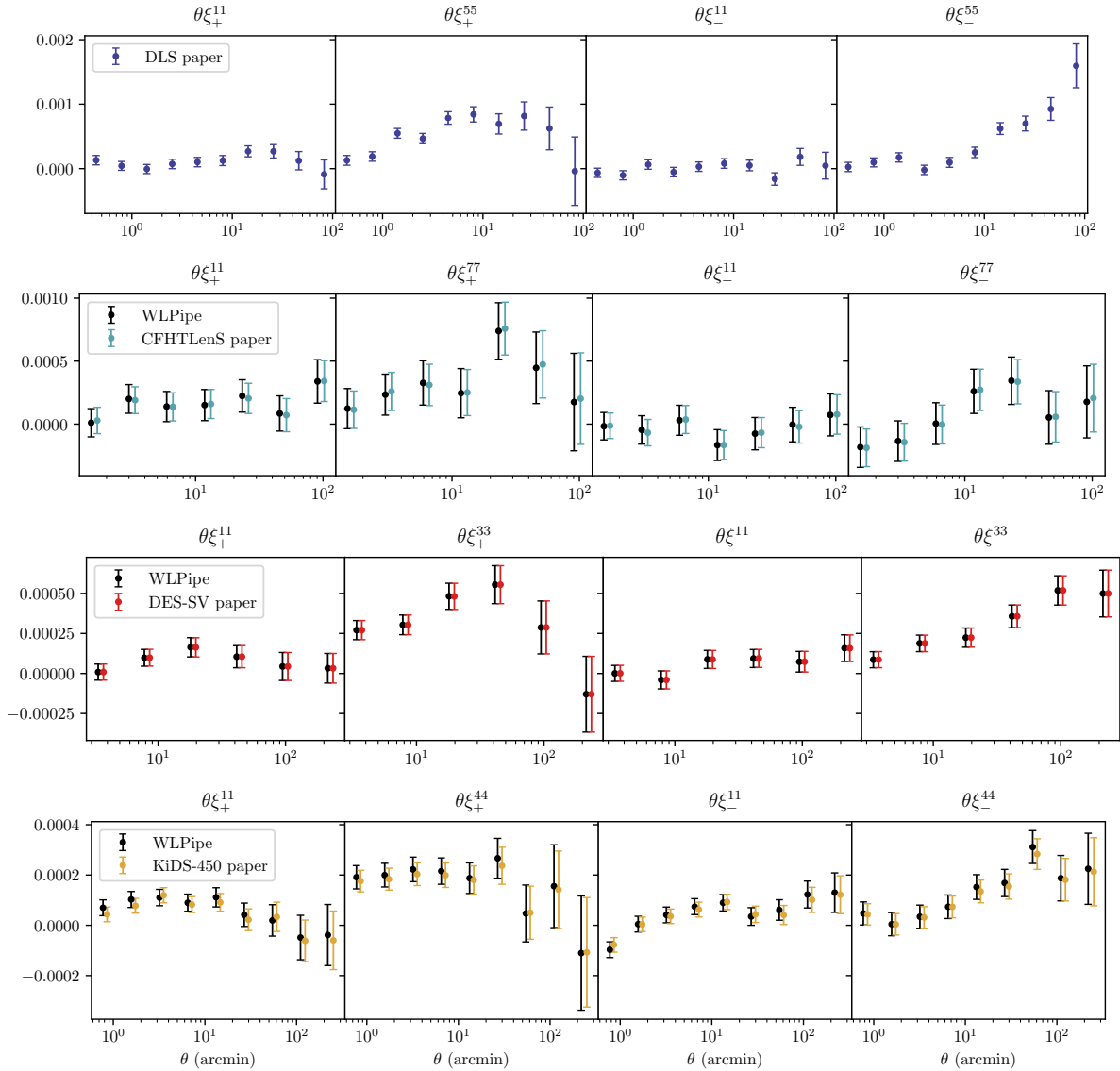
Figure 4. The 2-point function $\theta\xi_\pm(\theta)$ as measured by WLPIPE from the catalogs provided by the collaborations compared with the results obtained by the collaborations themselves. For visualization purpose we only show the auto-correlation functions for the lowest and the highest redshift bins, and the colored data points are slightly displaced from the black points. From left to right in each panel is ξ_+ for the lowest redshift bin, ξ_+ for the highest redshift bin, ξ_- for the lowest redshift bin, and ξ_- for the highest redshift bin. From top to bottom are the four surveys: DLS, CFHTLens, DES-SV, and KiDS-450. Since the catalogs from DLS are not public, only the collaboration 2-point functions are shown in the top panel. We also note that the difference in the angular binning discussed in Sec. 5.1 is not shown in this plot, but explained more clearly in Appendix A.

Sec. 5.2, Sec. 5.3 and Sec. 5.4, we investigate individually the effect of changing the covariance estimation, the scale cuts, and the priors on cosmological parameters, and intrinsic alignment treatment. In Sec. 5.5 we unify the analysis choices and reexamine the comparison metrics. In Sec. 5.6 we discuss how the definition of S_8 may affect the comparison between the surveys.

Throughout, we will also use the term *Nominal Baseline* to refer to the nominal analysis results that each collaboration uses as their most representative cosmological constraints, which for the case of CFHTLenS and KiDS-450 can be slightly different from the *Published Baseline* in terms of the treatment of systematic effects.

5.1 *Baseline:* Reproducing Literature Results

The most basic test is the comparison of the literature results with the WLPIPE's measurements using the same catalogs under the same assumptions.

First, we examine the intermediate output of the measured $\xi_\pm$ functions. Fig. 4 shows $\xi_+(\theta)$ and $\xi_-(\theta)$ produced by WLPIPE using the same binning and angular scales chosen by the collaborations, overlaid on top of results obtained by the collaborations for comparison. We find excellent agreement in all cases for the values of $\xi_\pm$[12]. Note that for DLS, CFHTLenS and KiDS-450,

---

[12] There is a small discrepancy between the published results and the WLPIPE measurements for KiDS-450 due to the fact that we have used the
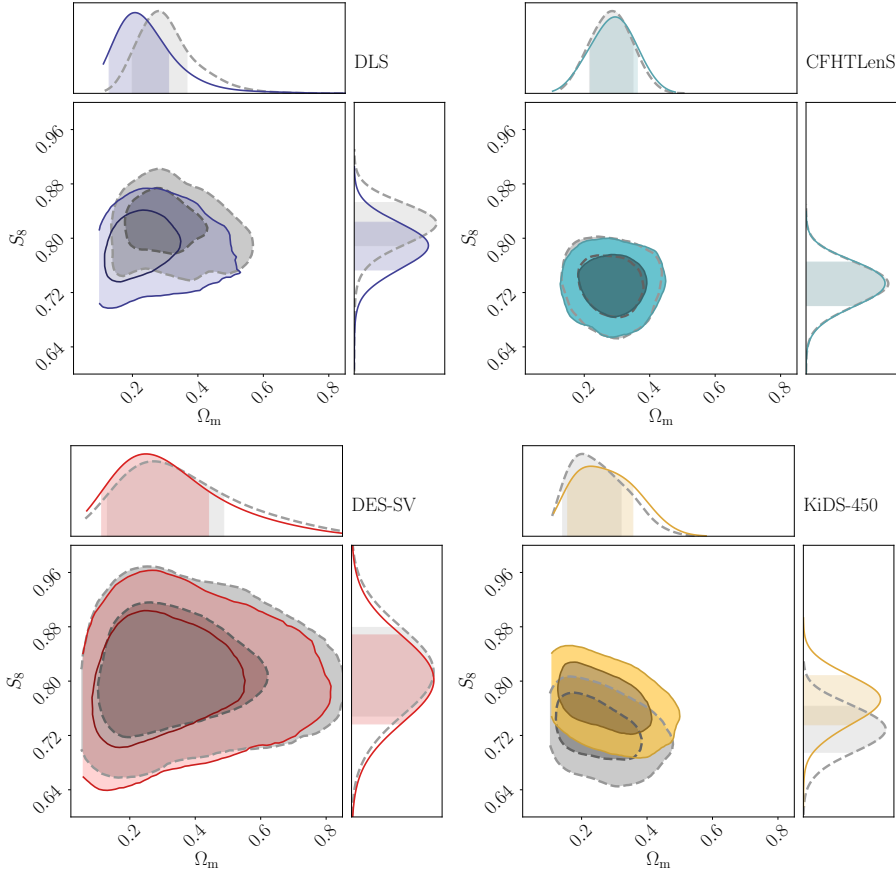
Figure 5. Results from reproducing published results with WLPipe. Here we compare the marginalized constraints on Ω_m and $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ obtained from WLPipe (*Baseline*, solid colored contours) to those obtained by the collaborations (*Published Baseline*, dashed grey contours) for four different experiments. One can see clear shifts in the DLS and KiDS-450 contours. Sec. 5.1 describes all the factors that drive these discrepancies, and any accidental cancellations of effects for the other surveys.

the angular values for each data point assigned by WLPipe differ from the paper-provided data vectors. This, as we discussed in Sec. 4.1, is due to the fact that those paper-provided data vectors used the center of each angular bin instead of the area-weighted center. We show how this propagates into a bias in the cosmological constraints in Appendix A.

Fig. 5 shows the constraints obtained from WLPipe for each experiment compared with those obtained by the collaborations themselves using the same binning, parameters, priors, and covariance matrices used to obtain the published results. In doing this we aim to reproduce the published results. However, Fig. 5 shows that there are differences between the published results and the WLPipe results, which we discuss in detail in the following subsections. The CosmoSIS configuration files and data files for these *Baseline* results are publicly available (Chang 2018).

5.1.1 DLS

From the upper left panel of Fig. 5, we see that the *Published Baseline* constraints from DLS are about 0.7σ higher in S_8 and 0.5σ higher in Ω_m than the *Baseline* constraints obtained via WLPipe.

per-object per-patch multiplicative bias correction instead of a constant for each tomographic bin used in Hildebrandt et al. (2017). We have checked that this does not affect the rest of the analysis.

Differences in angular binning cannot be an issue here, since we are using the collaboration-computed $\xi_\pm$. Two differences in the analysis explain the offset: First, the nonlinear power spectrum used in the original DLS analysis of Jee et al. (2016) comes from an older version of Halofit (Smith et al. 2003), while in CosmoSIS we use the nonlinear power spectrum of Takahashi et al. (2012). As shown in MacCrann et al. (2015) and also discussed in Sec. 6.3 of Jee et al. (2016), switching from the Smith et al. (2003) model to the Takahashi et al. (2012) model causes the inferred σ_8 to be lowered by ~ 0.02 at $\Omega_m = 0.3$. Second, as we have not implemented the particular IA model used in Jee et al. (2016) in WLPipe, we are assuming no IA in the WLPipe case. According to Figure 12 of Jee et al. (2016), this results in a ~ 0.02 lower Ω_m (with approximately the same S_8). Accounting for these two factors brings the two contours to better agreement – where the WLPipe reproduction gives a slightly lower Ω_m, but almost exactly the same S_8 compared to the published results.

5.1.2 CFHTLenS

From the upper right panel of Fig. 5, we see that the published constraints from CFHTLenS are consistent with WLPipe in both the Ω_m and S_8 directions. We note that we have chosen to compare the "fiducial" chain in Joudaki et al. (2017a), which does not include IA, baryons, photo-z uncertainties or shear calibration un-

certainties. Three factors need to be accounted for here: First, the angular values used in the paper-provided chains (the center of the bin) are different from that in the WLPIPE chain [Eq. (5)]. As we show in Appendix A, using the area-weighted angular values would shift the contours up by about 0.4σ. Second, whereas COSMOSIS uses the Takahashi et al. (2012) model in HALOFIT, Joudaki et al. (2017a) used the slightly more accurate HMCODE (Mead et al. 2016) for the nonlinear power spectrum. As shown in Figure 10 of Joudaki et al. (2017a), the HMCODE version used at that time moves the contour higher in S_8 by about 0.4σ compared to HALOFIT. These first two effects cancel, bringing the paper-provided chains and the WLPIPE reproduction to perfect agreement. The final difference in our approaches is more subtle. As we noted in Sec. 4.3, CFHTLenS and KiDS-450 uses COSMOMC, which does not sample h directly. Instead, it samples a wide flat prior in θ_{MC} (which is connected to h) and imposes the h priors after the fact. This means that the real h prior in the paper-provided chains is not exactly flat. This difference has been found to be small (Hildebrandt et al. 2017; Troxel et al. 2018).

In Sec 5.1-5.4, we compare with the "fiducial" case in Joudaki et al. (2017a) for simplicity. This assumes no systematic uncertainties, which according to Fig. 12 of Joudaki et al. (2017a) and Fig. 13, is close to the MID case in Joudaki et al. (2017a) (S_8 is lower by 0.1σ). We do incorporate systematic uncertainties for CFHTLenS in Sec. 5.5 based on Kilbinger et al. (2017); Choi et al. (2016).

5.1.3   DES-SV

We expect the WLPIPE reproduction of the DES-SV *Published Baseline* results to be perfect up to noise in the sampling, since the analysis pipeline is almost identical in the two analyses (WLPIPE uses slightly updated versions of TreeCorr, COSMOLIKE and COSMOSIS compared to that used in DES Collaboration et al. 2016). As shown in the lower left panel of Fig. 5, this is indeed the case – the two contours agree very well.

5.1.4   KiDS-450

From the lower right panel of Fig. 5, we see that the *Published Baseline* constraints from KiDS-450 agree with the *Baseline* constraints from WLPIPE in the Ω_m direction and are about 0.9σ higher in the S_8 direction. Several factors contribute to this discrepancy at different levels. First, the angular values used in the paper-provided chains (the center of the bin) are different from those in the WLPIPE chain [Eq. (5)]. Changing the bin values shifts the paper-provided chains up by about 0.4σ as shown in Fig. A1. Second, similar to CFHTLenS, the paper-provided chain uses HM-CODE for the nonlinear power spectrum while WLPIPE uses HALOFIT. However, while Joudaki et al. (2017a) used the original version of HMCODE (Mead 2015), a newer version of HM-CODE (Mead et al. 2016) was used in Hildebrandt et al. (2017). In this newer version, the fitting parameters were updated to allow for better fits when considering massive neutrino cosmologies, at the expense of slightly worse fits in standard ΛCDM. This newer version of HMCODE agrees more strongly with HALOFIT, and the resulting parameter constraints from KiDS-450 when using either prescription are almost identical (when excluding baryonic feedback). Third, similar to CFHTLenS, θ_{MC} is varied in the analysis while h is a derived parameter. Fourth, the covariance used in Hildebrandt et al. (2017) is designed to include the marginalization

Table 3. Comparison metrics for all pairs of surveys in the *Baseline* analysis case: WLPIPE chains that are designed to match the published analyses, or the *Published Baseline* case. For the S_8 values, we list the mean and the 16% and 84% confidence intervals. We note that here we have used the different analysis choices based on each of the collaborations, so these metrics are not on equal footing. Later in Table 6 we show similar metrics that *can* be compared directly.

	(1) DLS	(2) CFHTLenS	(3) DES-SV	KiDS-450
S_8	$0.80^{+0.032}_{-0.032}$	$0.73^{+0.028}_{-0.028}$	$0.80^{+0.059}_{-0.058}$	$0.77^{+0.033}_{-0.034}$
S/N	21.5	22.7	10.6	16.3
χ^2/ν	334.8/235	417.6/275	26.9/30	122.4/124
p.t.e.	2.0×10^{-5}	6.0×10^{-8}	0.63	0.52
ΔS_8-(1)	–	1.5	0.10	0.56
ΔS_8-(2)	–	–	1.1	0.87
ΔS_8-(3)	–	–	–	0.48

over multiplicative bias m, but was not implemented correctly (see also Fig. B1 and Troxel et al. 2018). This moves the S_8 constraints up by about 0.5σ. Finally, as described in Sec. 2.4, the photo-z uncertainties are incorporated differently in WLPIPE compared to Hildebrandt et al. (2017). We have checked, however, that this does not generate any noticeable effect in the cosmological constraints.

In summary, we are able to reproduce the Hildebrandt et al. (2017) results in both Ω_m and S_8 when considering these factors. We note here that the fiducial analysis of Hildebrandt et al. (2017) includes modeling of the baryonic effects on small scales whereas we do not here. As a result we compare with their DIR chain, which as shown in Fig. 8 of Hildebrandt et al. (2017) and Fig. 13, gives a S_8 value 0.3σ lower than the *Nominal Baseline* case. Later when we unify the analysis choices, since we make much more conservative scale cuts, we do not expect the effect of baryons to be important.

5.1.5   Comparison of all four surveys

The right panel of Fig. 6 shows the *Baseline* results from the four experiments using WLPIPE in one plot, i.e., we overlay the colored contours in Fig. 5 together. We note that here we have used the different analysis choices based on each of the collaborations, therefore the four contours cannot be compared on an equal footing. In this picture, we find good agreement between the four surveys in the $\Omega_m - S_8$ plane, with CFHTLenS slightly lower than the other three surveys. DES-SV has the largest contour (weakest constraining power), whereas the other three surveys have contours of similar sizes. The degeneracy directions of the four surveys are somewhat different, as expected from the different redshift ranges they probe. For comparison, we also show in the left panel of Fig. 6 the *Published Baseline* results from the corresponding survey-provided chains, or the four grey dashed contours in Fig. 5. The main difference from Fig. 5 is (1) the shifting of the KiDS-450 contours in the S_8 direction, which comes from the change in the angular bin values and the covariance, as we discussed in Sec. 5.1.4 above, and (2) the DLS contours shifted to lower S_8 due to the change in the nonlinear power spectrum and the IA model, as we discussed in Sec. 5.1.1 above. This can also be seen more clearly comparing the *Published Baseline* and *Baseline* cases in Fig. 13.

We list the comparison metrics (as described in Sec. 4.5) for all the surveys as well as combinations of survey pairs for the chains in the *Baseline* case in Table 3. First, looking at the S/N, we notice
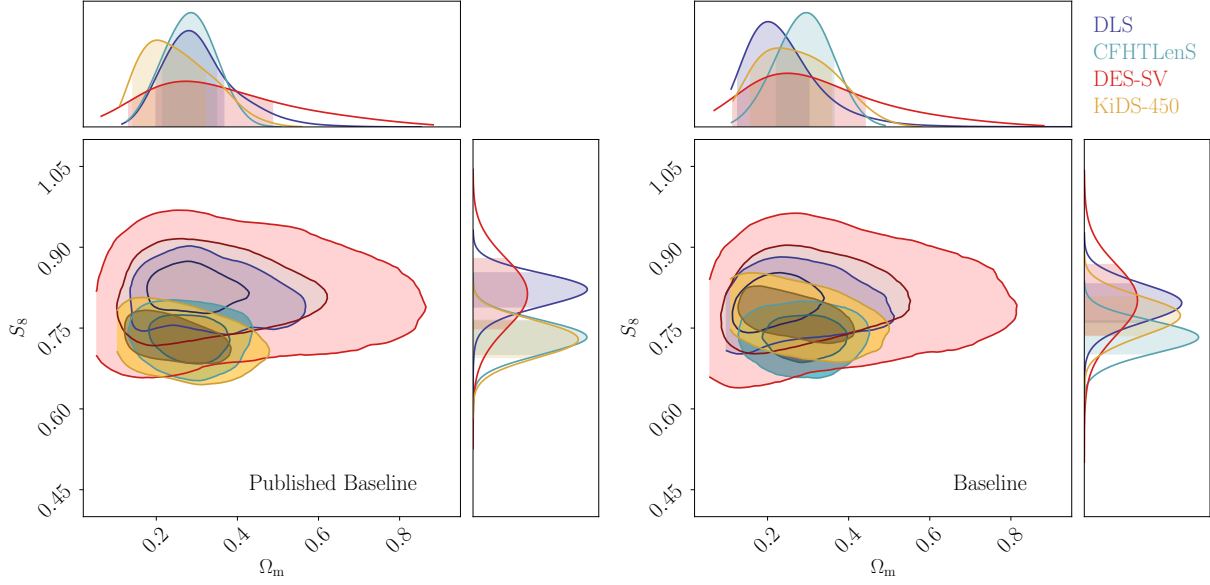
Figure 6. Here we compare the constraints of the four surveys from the published results and the WLPIPE reanalysis. We show the marginalized constraints on $\Omega_{\rm m}$ and $S_8 \equiv \sigma_8(\Omega_{\rm m}/0.3)^{0.5}$ from the paper-provided chains (the *Published Baseline* case, left panel) and from WLPIPE in the *Baseline* case. Note that compared to the *Published Nominal* results, here the KiDS-450 contours do not include baryonic effects, while the CFHTLenS contours do not include any systematic uncertainties.

Table 4. Comparison metrics for all pairs of surveys in the *Published Baseline* analysis case: constraints from the individual collaborations that we choose as baseline to reproduce. For the S_8 values, we list the mean and the 16% and 84% confidence intervals. For CFHTLenS and KiDS-450, these are different from the *Published Nominal* analysis case: constraints from the individual collaborations that can be viewed as the representative results.

	(1) DLS	(2) CFHTLenS	(3) DES-SV	(4) KiDS-450
S_8	$0.82^{+0.030}_{-0.030}$	$0.73^{+0.030}_{-0.030}$	$0.81^{+0.059}_{-0.058}$	$0.73^{+0.033}_{-0.032}$
ΔS_8-(1)	–	2.1	0.076	2.1
ΔS_8-(2)	–	–	1.2	0.087
ΔS_8-(3)	–	–	–	1.3

Table 5. S_8 constraints, S/N and goodness of fit when we change one analysis choice at a time in the analysis pipeline from the *Baseline* case (see Table 3). For the S_8 values, we list the mean and the 16% and 84% confidence intervals. The sections of this table correspond to discussions in Sec. 5.2, Sec. 5.3 and Sec. 5.4.

	(1) DLS	(2) CFHTLenS	(3) DES-SV	(4) KiDS-450
	Gaussian covariance matrix (Sec. 5.2)			
S_8	$0.84^{+0.030}_{-0.030}$	$0.74^{+0.024}_{-0.025}$	$0.83^{+0.052}_{-0.050}$	$0.77^{+0.030}_{-0.030}$
S/N	26.0	22.2	12.7	20.4
χ^2/ν	412.5/235	344.3/275	34.6/30	133.0/124
p.t.e.	7.0×10^{-12}	0.0028	0.26	0.27
	Conservative scale cuts (Sec. 5.3)			
S_8	$0.93^{+0.050}_{-0.050}$	$0.73^{+0.052}_{-0.050}$	$0.80^{+0.068}_{-0.069}$	$0.75^{+0.055}_{-0.055}$
S/N	15.4	16.6	10.0	10.5
χ^2/ν	112.1/89	228.3/132	28.4/25	62.8/56
p.t.e.	0.050	4.0×10^{-7}	0.29	0.24
	DES-SV priors (Sec. 5.4)			
S_8	$0.85^{+0.042}_{-0.042}$	$0.66^{+0.052}_{-0.052}$	$0.80^{+0.059}_{-0.058}$	$0.76^{+0.038}_{-0.038}$
χ^2/ν	319.5/235	412.2/275	26.9/30	121.5/124
p.t.e.	2.0×10^{-4}	1.6×10^{-7}	0.63	0.55
	KiDS-450 priors (Sec. 5.4)			
S_8	$0.82^{+0.033}_{-0.033}$	$0.68^{+0.039}_{-0.039}$	$0.81^{+0.059}_{-0.059}$	$0.77^{+0.033}_{-0.033}$
χ^2/ν	323.6/235	412.5/275	27.0/30	122.2/124
p.t.e.	1.1×10^{-4}	1.5×10^{-7}	0.63	0.53

that in the data configuration used in the individual surveys, the raw statistical power of the measurement is similar for DLS and CFHTLenS, while DES-SV is about half the S/N and KiDS-450 is in between. One interesting observation is that DLS achieves the high S/N even with a significantly smaller area – this highlights the power of having high-redshift data. A slightly worrying point is that the goodness-of-fits for DLS and CFHTLenS are quite low. For the pair-wise ΔS_8, we find trends reflecting what is seen from the figures – all four surveys are broadly consistent with Table 3 showing some low-level discrepancies (1.5σ) in S_8 between CFHTLenS and DLS.

For the *Published Baseline* chains, we list the S_8 constraints and ΔS_8 values in Table 4. We do not list the goodness-of-fit here since they are not all available in the papers, and are not directly comparable with the values in Table 3. We just quote two numbers that available: in Joudaki et al. (2017a), the reduced χ^2 for the fiducial CFHTLenS analysis best-fit is 1.5, whereas in Hildebrandt et al. (2017), the reduced χ^2 for the fiducial KiDS-450 analysis

best-fit is 1.3. In Troxel et al. (2018), it was shown that the reduced χ^2 for the fiducial KiDS-450 improves to 1.0 when accounting for the survey geometry.
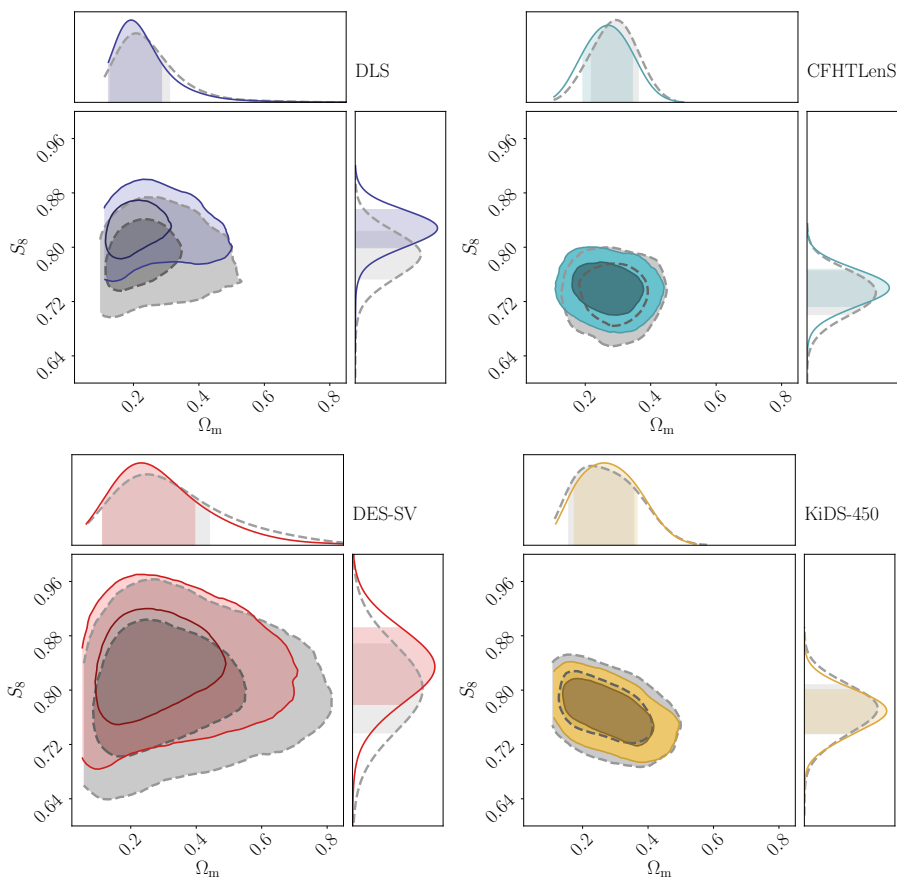
Figure 7. Effect of the covariance estimation – marginalized constraints on Ω_m and $S_8 \equiv \sigma_8 (\Omega_m/0.3)^{0.5}$ obtained from WLP IPE using the survey-provided covariance (grey dashed) for four different experiments and the COSMOLIKE Gaussian analytic covariance (colored solid). We can see a shift in the contours for DLS, CFHTLenS and DES-SV.

5.2 Effect of the Covariance Matrix

Now we investigate the effect of the covariance matrix estimation. As discussed in Sec. 4.2, the four surveys have different approaches to covariance estimation. We eliminate these differences by generating a Gaussian analytical COSMOLIKE covariance matrix for each survey.

Fig. 7 shows the changes in the contours in the four experiments when analytic covariance matrices are used in place of those provided by the collaborations. The corresponding comparison metrics are listed in Table 5. We notice a shifts of the contours in the S_8 constraints for some of the surveys. Overall, the Gaussian analytic covariance leads to slightly tighter constraints compared to covariance matrices estimated from simulations. This could be partially due to the fact that we have not accounted for the non-Gaussian piece of the analytic covariance.

For DLS, we see a significant shift in the mean of the constraints towards higher S_8 values; DES-SV and CFHTLenS also show some shifts in S_8, but less significant. We note that, since the data vector is noisy, we do not expect the contours to agree exactly. However, we believe the shift for DLS is more than what is expected from statistical fluctuation. The DLS field is much smaller and contains a lower level of shape noise compared to the other surveys. In addition, one of the fields contains a galaxy cluster. These factors mean that the covariance is challenging to model and the simple Gaussian covariance used here may not be a good approximation for the dataset. It is possible that neither the

survey-provided covariance nor the Gaussian COSMOLIKE covariance from WLP IPE captures these complications. We also note that for the three cases where simulation covariance is used, DES-SV has the smallest Hartlap factor ($H_{DLS} = 0.88$, $H_{CFHTLenS} = 0.86$, $H_{DES-SV} = 0.7$). This means that the inverse of the simulation covariance in DES-SV is expected to be noisier (but unbiased) compared to the other two simulation covariances (Dodelson & Schneider 2013).

Finally, it is also worth noting that since the survey-provided covariance from KiDS-450 is also an analytic covariance matrix, the agreement between the dashed and the solid contours in the bottom right of Fig. 7 is a good check on the analytic calculation for the covariance. We have checked that the slightly smaller contours from WLP IPE is partially reflecting the difference between the Gaussian and non-Gaussian covariance.

5.3 Effect of Scale Cuts

In this section we investigate the effect of scale cuts. Following Sec. 4.4, we choose to match all scale cuts to the most conservative scale cuts in the four datasets ($R_{min,+} > 1.3$ Mpc and $R_{min,-} > 11.4$ Mpc, see Eq. (12)). The results are shown in Fig. 8, with the corresponding metrics listed in Table 5. The exact cuts used in each bin are tabulated in Appendix C, Table C1. In all these tests, everything else in the analysis stays the same as the *Baseline* case in Sec. 5.1.

In Fig. 8, the first thing that draws the eye is the DLS contours,
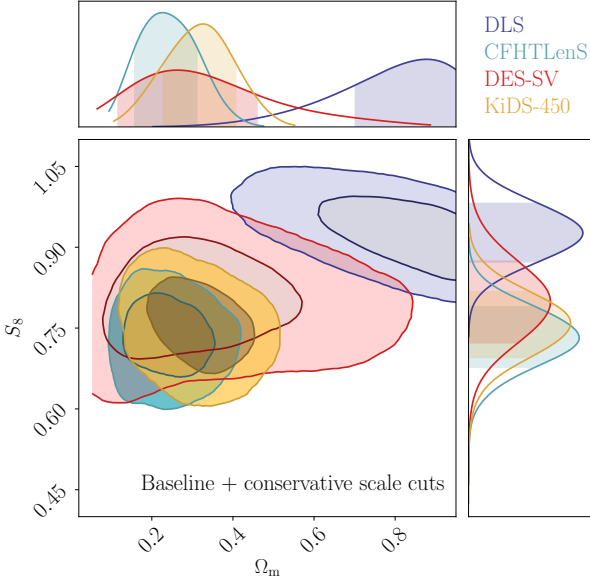
Figure 8. Effect of scale cuts compared to *Baseline* (right panel of Fig. 6) – we show the marginalized constraints for Ω_m and $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$ when small scale data are removed from the fit (requiring $R_\mathrm{min,+} > 1.3$ Mpc and $R_\mathrm{min,-} > 11.4$ Mpc). DLS shifts to large Ω_m and S_8 values, while the other surveys show enlarged contours compared to *Baseline*. We note that the four contours should not be compared directly here, as the analysis choices are not unified.

which shift to very large Ω_m values, as well as a higher S_8. All the other surveys appear consistent with the original case in Fig. 6, but with looser constraints due to the fact that we have removed information.

We note that the goodness-of-fit for DLS improved significantly when applying the conservative scale cuts compared to the *Baseline* case. After a more careful look at the DLS measurements, it appears that the small-scale data points for ξ_- is the source of the contour shift – those data points prefer a lower amplitude compared to the rest of the data points. Therefore, when applying the conservative scale cuts, the model amplitude increases (so does S_8), and the goodness-of-fit improves. This could also be a hint that the small-scale covariance is underestimated, as already discussed in Sec. 5.2, that the characteristics of the DLS data makes it difficult to model the covariance. We note that some of these issues were discussed in Jee et al. (2013) and Jee et al. (2016), and a similar trend in S_8 was seen in Fig.13 of Jee et al. (2016). Here we caution that since the DLS contours are far from $\Omega_\mathrm{m} = 0.3$ and clipped by the priors ($\Omega_\mathrm{m} < 1$), the S_8 values quoted are not very meaningful.

5.4 Impact of Cosmological Priors and IA Treatment

Next, we consider the impact of different cosmological priors and IA treatments. To address this, we impose identical priors on all surveys, first using those from DES-SV and then from KiDS-450 (see Table 2) since they roughly represent the two approaches of handling the parameters: DES-SV has priors that are relatively conservative, and in the parametrization of $[\Omega_\mathrm{m}, \Omega_\mathrm{b}, h, \sigma_8, n_s]$, whereas KiDS-450 has more restrictive priors and uses the parametrization $[\Omega_c h^2, \Omega_\mathrm{b} h^2, h, \ln(10^{10}A_s), n_s]$. We moreover allow for intrinsic alignments in the case of CFHTLenS and DLS. For all surveys, we consider either the IA amplitude prior $-5 < A_\mathrm{IA} < 5$ used by DES-SV or the prior $-6 < A_\mathrm{IA} < 6$ used by KiDS-450. Note that aside
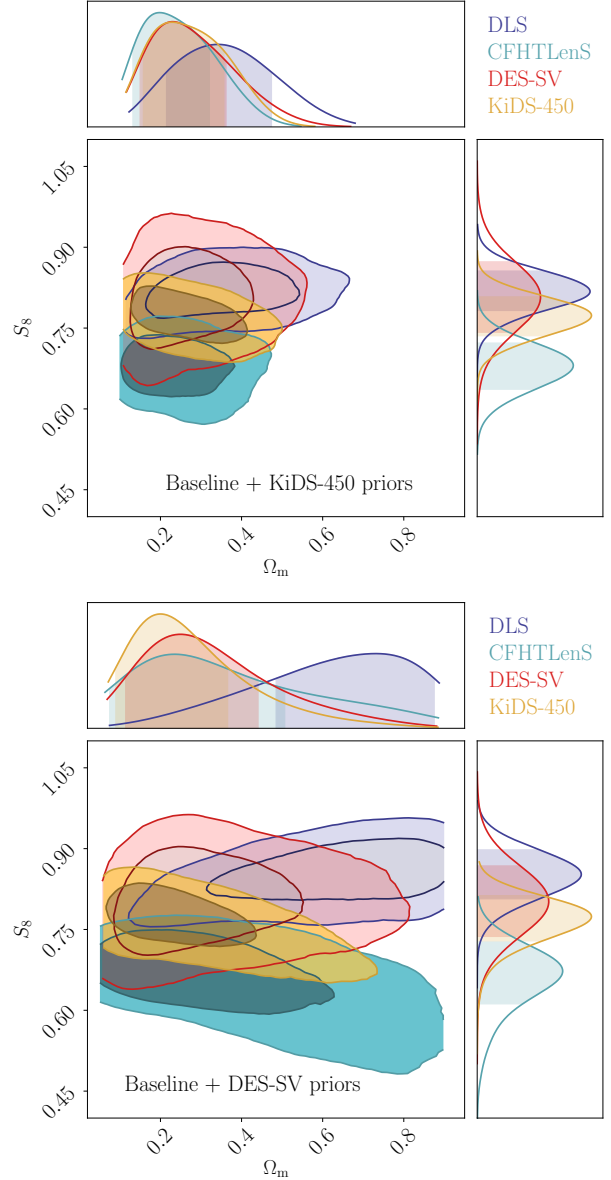
Figure 9. Impact of cosmological priors and IA treatments compared to *Baseline* (right panel of Fig. 6) – we show the marginalized constraints for Ω_m and $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$ when unifying the priors on the cosmological and IA parameters. We first unify to the KiDS-450 priors (top), then to the DES-SV priors (bottom). The constraints in the Ω_m direction is heavily affected by the priors, while in the S_8 direction, there is a larger effect for surveys with a strong degeneracy in the $\Omega_\mathrm{m} - S_8$ plane. We note that the four contours should not be compared directly here, as the analysis choices are not unified.

from these changes to the cosmological priors and IA treatments, we keep all other analysis choices the same as in the *Baseline* case of Sec. 5.1.

The two panels of Fig. 9 show the effect of unifying the cosmological priors and IA treatments from that chosen as *Baseline*, with the corresponding metrics listed in Table 5. Looking at CFHTLenS, DES-SV and KiDS-450, it is apparent that the constraints in the Ω_m direction are largely dominated by cosmological priors. Specifically, the prior on h, which is wider for DES-SV compared to KiDS-450, leads to large changes in the Ω_m posterior. The
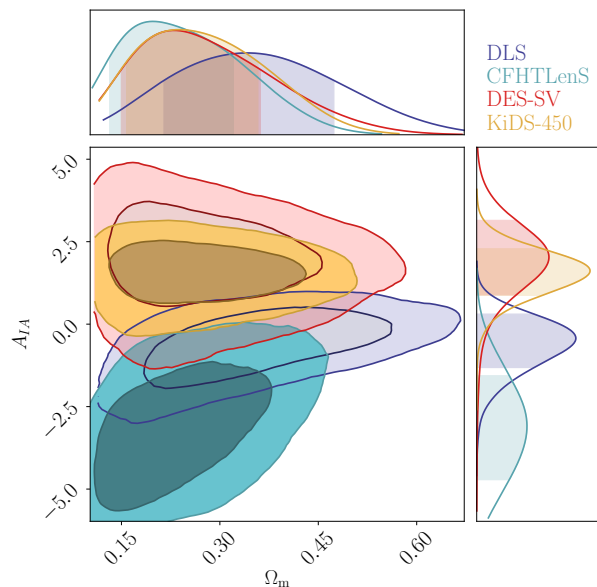
Figure 10. Constraints on IA amplitude – we show the marginalized constraints for Ω_m and A_{IA} when unifying the priors to the KiDS-450 priors (keeping all other analysis choices as in the *Baseline* case). We find a degeneracy between A_{IA} and Ω_m for DLS and CFHTLenS. We also find that both these surveys prefer negative IA amplitudes in this set-up.

constraints on S_8, on the other hand, are relatively robust to cosmological priors, consistent with previous findings (Kilbinger et al. 2013; Joudaki et al. 2017a). This again is showing that cosmic shear measurements for these four datasets are mainly constraining only the amplitude of the power spectrum and not the detailed shape of it. The uncertainty on S_8 decreases for CFHTLenS when moving to tighter cosmological priors, however, this is largely due to the fact that the S_8 definition here is not optimal for the CFHTLenS dataset. We will discuss this point in Sec. 5.6.

A few other effects of the cosmological priors and IA treatment are visible in Fig. 9. First, for DLS, when imposing the DES-SV priors, Ω_m moves to high values while S_8 remains roughly the same. When imposing KiDS-450 priors, the Ω_m constraints appear similar to the *Baseline* case. This behavior, together with what is shown in Sec. 5.3, suggests that the DLS constraints on Ω_m are sensitive to the scales used and the priors. For CFHTLenS, the S_8 constraints move to lower values using both DES-SV and KiDS-450 priors. This comes from the fact that compared to the *Baseline* case, here there is additional freedom in the IA amplitude. We examine the IA amplitude when using the KiDS-450 priors, as shown in Fig. 10, and find that the CFHTLenS favors a negative IA amplitude at the 2σ level. This, based on previous work in measurements of IA, suggests that we may be fitting to some systematic effects that appear to behave like IA (Kilbinger et al. 2017; Choi et al. 2016; van Uitert et al. 2018). This is consistent with Fig. 8 and Fig. 9 of Joudaki et al. (2017a), where they show that this negative IA shifts the S_8 constraints to lower values. There is also a similar (but less severe) trend in the DLS data.

5.5 Common Covariances, Angular Scale Cuts, Cosmological Priors, and IA Treatments

After investigating the individual effects in Sec. 5.2, Sec. 5.3 and Sec. 5.4, we now combine all of them and perform a uniform anal-

ysis on all four surveys. We study two cases, both using COSMO-LIKE Gaussian covariances, conservative scale cuts, the same IA treatments, and we use two sets of priors:

(i) KiDS-450 priors and
(ii) DES-SV priors.

As discussed in Sec. 5.1.2, in this subsection we incorporate the photo-z and shear calibration bias uncertainties for CFHTLenS. As summarized in Sec. 4.3 of Kilbinger et al. (2017), a number of improvements to CFHTLenS have been identified since the public release of the catalogues in 2013. Of importance to this study is the analysis by Choi et al. (2016) who showed that significant biases existed in the reported photo-z distributions, the result from Kuijken et al. (2015) that the CFHTLenS shear calibration corrections were in general underestimated and the finding by Fenech Conti et al. (2017) that the previously unexplored area of galaxy selection bias results in a few percent overestimation of the shear calibration correction. The conclusion of all these works was that any future analyses of CFHTLenS should include conservative systematic error terms to account for these effects. In this section, we therefore marginalize over an uncertainty in the mean redshift of each bin with zero-mean top-hat prior of full-width 0.2, and an uncertainty in the shear calibration correction zero-mean top-hat prior of full-width 0.1.

In Fig. 11 we show the comparison between the *Published Nominal* contours and case (i) listed above. The *Published Nominal* contours present the view one would have on the four cosmic shear surveys after reading the individual papers (Jee et al. 2016; Joudaki et al. 2017a; DES Collaboration et al. 2016; Hildebrandt et al. 2017), while the *Matched* contours present what the cosmological constraints are when analysed through a unified analysis framework. One can also compare with Fig. 6 to understand the nature of the different changes in the contours. The contours for (ii) are shown in Fig. 12. We choose to focus on (i) here as it is not as affected by the S_8 definition (see Sec. 5.6) as (ii). The four surveys in the right panel of Fig. 11 can now be compared on equal footing. The comparison metrics are shown in Table 6.

In general, we observe the same effects seen individually in Sec. 5.2, Sec. 5.3 and Sec. 5.4. But when put together, the discrepancies between the different surveys coming from the different effects accumulate and become larger. Looking at Fig. 11 and the ΔS_8 statistics in Table 6, we find that essentially none of the surveys have S_8 constraints that agree within 1σ and the extreme cases differ more than 3σ. If we look at the change in the S_8 constraints for the individual surveys from the left panel to the right panel, it is clear that the main effect is that DLS moves to larger S_8. CFHTLenS is consistent with the MIN and MAX case but not the MID case, which is expected given that the choice of IA models we use is the same as the MIN case, and that the MAX case has little constraining power. DES-SV and KiDS-450 stay roughly the same.

Next we turn to the other statistics in Table 6. We note that the signal-to-noise for the four datasets change slightly, but the relative power stays roughly the same, with DLS being the highest and DES-SV being the lowest. We note that the goodness-of-fit for DLS and CFHTLenS improved from the *Baseline* case but is still quite low. The largest ΔS_8 is about 3.4σ between DLS and CFHTLenS, which is also apparent from Fig. 11. Next we look at the BF statistic [Eq. (17)] between pairs of surveys. Here when evaluating the numerator in Eq. (17) for BF, we only require the cosmological parameters to be shared amongst the two experiments being compared and keep the IA amplitude, shear calibration parameter and photo-z uncertainty separate. We find that the message from the BF statis-
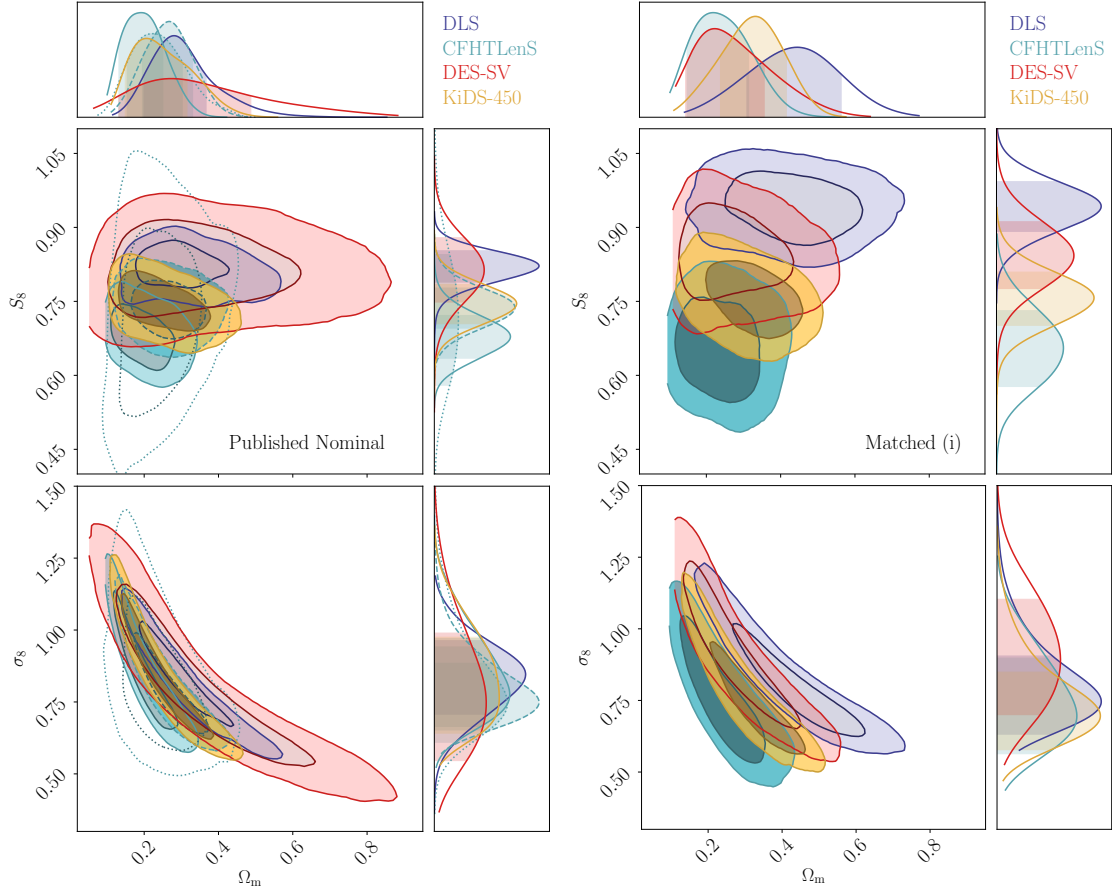
Figure 11. Final comparison of the cosmological constraints from the four surveys according to the literature (*Published Nominal*, left) and according to our unified analysis framework (right, *Matched*) – we show the marginalized constraints for Ω_m, $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ and σ_8 for the four cosmic shear surveys. In the right panel, we use Gaussian analytic covariances, conservative scale cuts and the KiDS-450 priors. We note that for the CFHTLenS *Published Nominal* constraints, we show all three settings MIN (solid), MID (dashed) and MAX (dotted) in Joudaki et al. (2017a).
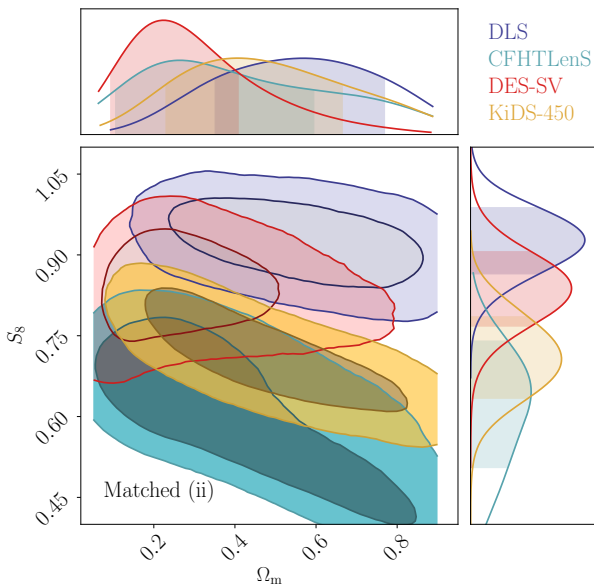
Figure 12. Same as the upper right panel of Fig. 11, but now using DES-SV priors.

Table 6. Comparison metrics corresponding to the right panel of Fig. 11. That is, all analysis choices matched: Gaussian COSMOLIKE covariance matrix, conservative scale cuts, same IA treatments, and KiDS-450 cosmological priors. For the S_8 values, we list the mean and the 16% and 84% confidence intervals.

	(1) DLS	(2) CFHTLenS	(3) DES-SV	(4) KiDS-450
S_8	$0.94^{+0.046}_{-0.045}$	$0.66^{+0.071}_{-0.070}$	$0.84^{+0.062}_{-0.061}$	$0.76^{+0.048}_{-0.049}$
S/N	17.4	15.1	11.6	12.1
χ^2/ν	137.8/89	176.3/132	32.7/26	71.5/56
p.t.e.	7.0×10^{-4}	0.0060	0.17	0.079
ΔS_8-(1)	–	3.4	1.3	2.9
ΔS_8-(2)	–	–	2.0	1.1
ΔS_8-(3)	–	–	–	1.2
BF-(1)	–	-1.1	1.6	-0.50
BF-(2)	–	–	0.70	1.3
BF-(3)	–	–	–	1.1

tics is similar to that captured by the ΔS_8 metric in this case, though the message of consistency/inconsistency is somewhat weaker – the

Figure 13. Summary of $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$ constraints from the four cosmic shear surveys studied in this work. Each bar shows the 16% and 84% confidence interval of the S_8 constraints. The plot is divided by dashed horizontal lines into four sections. The first section shows chains provided by the respective collaborations. The second section shows constraints derived from WLPIPE, but are not fully matched in all analysis steps, ergo not directly comparable. The third section shows constraints derived from WLPIPE and have analysis steps matched, so can be compared directly. The last section shows the constraints from combining the two consistent surveys DES-SV and KiDS-450, the *Planck* satellite (Planck Collaboration et al. 2018) and DES Y1 cosmic shear (Troxel et al. 2017). The naming conventions are explained in the table below the figure. We note that for the CFHTLenS *Published Nominal* constraints, we show all three settings MIN (top), MID (middle) and MAX (bottom) in Joudaki et al. (2017a).

only BF value that fails the requirement of combining is DLS and CFHTLenS.

We now combine the two surveys, DES-SV and KiDS-450, under this matched setting. These two datasets are consistent under the same model assumption according the BF metric, and have reasonable goodness-of-fit values. The combined constraint of DES-SV and KiDS-450 is $S_8 = 0.79^{+0.042}_{-0.041}$. Compared to the cosmic shear results from the first year DES data (Troxel et al. 2017), $S_8 = 0.79^{+0.024}_{-0.026}$, we find excellent agreement. Compared to state-of-the-art CMB constraints (Planck Collaboration et al. 2018), $S_8 = 0.83^{+0.013}_{-0.013}$, we find reasonably consistent results with roughly 1σ lower S_8. These results are in good agreement with that found in Troxel et al. (2018).

5.6 A side note on the S_8 definition

As discussed briefly in Sec. 4.5, S_8 is defined as $\sigma_8(\Omega_m/0.3)^{\alpha}$, where α is designed to remove the degeneracy between σ_8 and Ω_m. That is, if α is chosen optimally, it characterizes the direction orthogonal to the $\Omega_m - \sigma_8$ contours. For datasets of different redshift distribution, the optimal α is different.

Throughout our analysis, we have fixed α to be 0.5, which

may not be optimal for all datasets. This implies that for datasets where α is further from 0.5, the projected uncertainties on $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$ are going to be slightly larger than if the optimal α were used, and that when comparing the different surveys they will tend towards being consistent. This can be seen clearly in Fig. 12, where the contours for CFHTLenS and KiDS-450 are tilted leading to larger uncertainties in the S_8 direction. The effect is much reduced when a tighter prior is imposed as in the right-hand panels of Fig. 11. Roughly, we find that with the DES-SV priors (corresponding to Fig. 12), the optimal α values are 0.56 (DLS), 0.71 (CFHTLenS), 0.51 (DES-SV) and 0.67 (KiDS-450). With the KiDS-450 priors (corresponding to the right panels of Fig. 11), the optimal α values are 0.52 (DLS), 0.52 (CFHTLenS), 0.52 (DES-SV) and 0.58 (KiDS-450). That is, we expect the discrepancies between the surveys in the single parameter that quantifies the amplitude to be sensitive to the priors and likely larger if an optimal α is used. On the other hand, the BF metric is insensitive to the choice of α so is a more robust measure of consistency.

6 SUMMARY AND DISCUSSION

In this paper we use a generic cosmic shear pipeline, WLPIPE, that takes in galaxy shear catalogs, calculates the two-point shear-shear correlation function via the software package TREECORR and the associated covariance matrix via the software package COSMO-LIKE, and finally carries out cosmological parameter inference via the software package COSMOSIS. The WLPIPE framework is constructed using the PEGASUS workflow engine, which takes care of data and code transfer between different computing resources seamlessly. This pipeline also serves as a prototype pipeline for future analysis pipelines in DESC.

We apply this pipeline to four existing cosmic shear surveys: the Deep Lens Survey (DLS), the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS), the Science Verification data from the Dark Energy Survey (DES-SV), and the 450 deg^2 release of the Kilo-Degree Survey (KiDS-450). The goal is to first reproduce the literature results, investigate the effect of different analysis choices adopted in each of the surveys, and finally unifying these different analysis choices in order to perform an apples-to-apples comparison of the survey results. In Fig. 13 we summarize the constraints on $S_8 \equiv \sigma_8(\Omega_{\mathrm{m}}/0.3)^{0.5}$ from all the cases studied in this work. We summarize our main findings below:

• We are able to reproduce a specific set of the published results from the four collaborations when following the same analysis choices to well within the uncertainties. In this *Baseline* case, the four surveys appear to be broadly consistent in terms of their constraints on S_8: $S_8 = 0.80^{+0.032}_{-0.032}$ (DLS), $0.73^{+0.028}_{-0.028}$ (CFHTLenS), $0.80^{+0.059}_{-0.058}$ (DES-SV), and $0.77^{+0.033}_{-0.034}$ (KiDS-450). However, we note that not all the model fits are good descriptions of the data – for DLS and CFHTLenS, the p-values for the fits are low, while for KiDS-450, the p-value is acceptable, but only after incorporating recent improvements for the covariance.

• In reproducing the published results, we investigate several issues in the published results: the angular bin values used in the data vector, the incorporation of nuisance parameters in the covariance, and the nonlinear power spectrum model and others. We find these details can shift the cosmological constraints by $\sim 0.5\sigma$. Most of these issues are known, but analyzing all four experiments systematically in this work gives a big picture view of how the four analyses agree and differ.

• Effect of the covariance matrix: constraints based on simulation-based covariances can be shifted from analytic covariances due to noise. In addition, the DLS covariance may not be well approximated by a Gaussian covariance due to the complexity of the data, the small area and the low shape noise.

• Effect of scale cuts: sensitivity of the cosmological constraints to scale cuts could indicate internal inconsistency of datasets or further issues with the covariance. It could also point to potential failures in the models at small scales (e.g. IA, nonlinear matter power spectrum, baryonic physics).

• Effect of priors: for parameters that are not constrained (e.g. Ω_{m}), the priors have an effect on the constraints, but for parameters that are constrained (e.g. S_8), the effect of priors is smaller, but not negligible. A wide prior on the IA amplitude can absorb other sources of systematic issues, which could explain the slightly negative IA amplitude constrained by CFHTLenS.

• When unifying all analysis choices discussed above, the four surveys give the following constraints: we find $S_8 \equiv \sigma_8(\Omega_{\mathrm{m}}/0.3)^{0.5}$ to be $0.94^{+0.046}_{-0.045}$ (DLS), $0.66^{+0.071}_{-0.070}$ (CFHTLenS), $0.84^{+0.062}_{-0.061}$ (DES-SV) and $0.76^{+0.048}_{-0.049}$ (KiDS-450). Specifically, DLS moves to higher

S_8 while CFHTLenS moves to lower S_8 compared to the *Baseline*. The change in the DLS constraints is primarily due to the scale cuts and covariance, while the change in CFHTLenS is due to the change in the IA treatment, and could be an indication of residual issues in the photo-z estimation. The goodness-of-fit values for DLS and CFHTLenS improved but is still low.

• We calculate the ΔS_8 statistics and the Bayesian evidence ratio (BF) between each of the two surveys (when analysis choices are unified). The S_8 constraints from the two most discrepant cases (DLS and CFHTLenS) differ by 3.4σ. The S_8 constraints for DES-SV and KiDS-450 in the final matched analysis appear consistent with the *Baseline* analysis as well as with each other. They also seem to be robust to the various analysis choices tested. Together with the more reasonable goodness-of-fit values and IA constraints, this is an encouraging indication for the field given that DES-SV and KiDS-450 are the most recent work amongst the four surveys.

• Based on all the above information, we decide to combine the DES-SV and KiDS-450 datasets (based on the goodness-of-fit, IA constraints and consistency). The combined constraint is $S_8 = 0.79^{+0.042}_{-0.041}$, which is in agreement with both the cosmic shear constraints from the first year of DES data in Troxel et al. (2017), and the CMB constraints from Planck Collaboration et al. (2018).

Cosmic shear measurements hold great promise in terms of the constraining power in cosmology. In order to fully exploit this power in upcoming and future cosmic shear surveys (DES, KiDS, HSC, Euclid, LSST, WFIRST), it is important to learn from the experiences accumulated over the past years in the community across the different collaborations and datasets. We have demonstrated that a number of analysis choices can result in significant changes in the cosmological constraints and should therefore be treated with care for the future analyses.

REFERENCES

Abell P. A., et al., 2009
Aihara H., et al., 2017, preprint, (arXiv:1702.08449)
Babuji Y., Chard K., Foster I., Katz D., Wilde M., Woodard A., Wozniak J., 2018, Proceedings of the International Workshop on Science Gateways
Bacon D. J., Refregier A. R., Ellis R. S., 2000, MNRAS, 318, 625
Bartelmann M., Schneider P., 2001, Physics Reports, 340, 291
Becker M. R., et al., 2016, Phys. Rev., D94, 022002
Benítez N., 2000, ApJ, 536, 571
Bonnett C., 2013, preprint, (arXiv:1312.1287)
Bonnett C., et al., 2016, Phys.Rev.D, 94, 042005
Bridle S., King L., 2007, New Journal of Physics, 9, 444
Chang C., 2018, A Unified Analysis of Four Cosmic Shear Surveys, doi:10.5281/zenodo.1404447, https://doi.org/10.5281/zenodo.1404447
Choi A., et al., 2016, MNRAS, 463, 3737
Coil A. L., et al., 2011, ApJ, 741, 8
Cooray A., Hu W., 2001, ApJ, 554, 56
DES Collaboration Abbott T. M. C., et al., 2016, Phys. Rev., D94, 022001
DES Collaboration et al., 2017, preprint, (arXiv:1708.01530)
Deelman E., et al., 2015, Future Generation Computer Systems, 46, 17
Dodelson S., Schneider M. D., 2013, Phys.Rev.D, 88, 063537
Einstein A., 1936, Science, 84, 506
Erben T., et al., 2013, Mon. Not. Roy. Astron. Soc., 433, 2545
Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, MNRAS, 467, 1627
Feroz F., Hobson M. P., Bridges M., 2009, Mon. Not. Roy. Astron. Soc., 398, 1601
Flaugher B., 2005, International Journal of Modern Physics A, 20, 3121
Flaugher B., et al., 2015, preprint, (arXiv:1504.02900)

Freedman W. L., 2017, Nature Astronomy, 1, 0121

Fu L., et al., 2008, A&A, 479, 9

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

Heymans C., et al., 2006, MNRAS, 368, 1323

Heymans C., et al., 2012, Mon. Not. Roy. Astron. Soc., 427, 146

Heymans C., et al., 2013, MNRAS, 432, 2433

Hildebrandt H., et al., 2012, MNRAS, 421, 2355

Hildebrandt H., et al., 2017, MNRAS, 465, 1454

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

Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101

Ivezić v., et al., 2008

Jain B., Seljak U., 1997, ApJ, 484, 560

Jarvis M., et al., 2016, Mon. Not. Roy. Astron. Soc., 460, 2245

Jee M. J., Tyson J. A., 2011, PASP, 123, 596

Jee M. J., Tyson J. A., Schneider M. D., Wittman D., Schmidt S., Hilbert S., 2013, Astrophys. J., 765, 74

Jee M. J., Tyson J. A., Hilbert S., Schneider M. D., Schmidt S., Wittman D., 2016, ApJ, 824, 77

Jeffreys H., 1961, Theory of Probability, third edn. Oxford, Oxford, England

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

Joudaki S., et al., 2017a, MNRAS, 465, 2033

Joudaki S., et al., 2017b, MNRAS, 471, 1259

Joudaki S., et al., 2018, MNRAS, 474, 4894

Kaiser N., Wilson G., Luppino G. A., 2000, ArXiv Astrophysics e-prints,

Kilbinger M., et al., 2013, Mon. Not. Roy. Astron. Soc., 430, 2200

Kilbinger M., et al., 2017, MNRAS, 472, 2126

Krause E., Eifler T., 2016

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

Krause E., et al., 2017, preprint, (arXiv:1706.09359)

Kuijken K., et al., 2015, MNRAS, 454, 3500

Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118

Limber D. N., 1953, ApJ, 117, 134

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

MacCrann N., Zuntz J., Bridle S., Jain B., Becker M. R., 2015, MNRAS, 451, 2877

Mandelbaum R., 2017, preprint, (arXiv:1710.03235)

Mandelbaum R., et al., 2015, MNRAS, 450, 2963

Marshall P., Rajguru N., Slosar A., 2006, Phys.Rev.D, 73, 067302

Mead A., 2015, HMcode: Halo-model matter power spectrum computation, Astrophysics Source Code Library (ascl:1508.001)

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

Mead A. J., Heymans C., Lombriser L., Peacock J. A., Steele O. I., Winther H. A., 2016, MNRAS, 459, 1468

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

Miller L., et al., 2013, MNRAS, 429, 2858

Muller G. P., Reed R., Armandroff T., Boroson T. A., Jacoby G. H., 1998, in D'Odorico S., ed., SPIE Vol. 3355, Optical Astronomical Instrumentation. pp 577–585, doi:10.1117/12.316833

Planck Collaboration et al., 2018, preprint, (arXiv:1807.06209)

Raveri M., Hu W., 2018, preprint, (arXiv:1806.04649)

Sato M., Hamana T., Takahashi R., Takada M., Yoshida N., Matsubara T., Sugiyama N., 2009, ApJ, 701, 945

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

Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132

Sheldon E. S., 2014, MNRAS, 444, L25

Smith R. E., et al., 2003, MNRAS, 341, 1311

Takada M., Hu W., 2013, Phys.Rev.D, 87, 123504

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152

Troxel M. A., et al., 2017, preprint, (arXiv:1708.01538)

Troxel M. A., et al., 2018, preprint, (arXiv:1804.10663)

Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, Nature, 405, 143

Zuntz J., et al., 2015, Astron. Comput., 12, 45

de Jong J. T. A., et al., 2015, A&A, 582, A62

van Uitert E., et al., 2018, MNRAS, 476, 4662

APPENDIX A: EFFECT OF ANGULAR VALUES

As discussed in Sec. 4.1, out of the four surveys, three of them did not use the weighted angular values in the data vector [Eq. (5)]. Correcting for this, the DLS angular values changed from [0.40, 0.72, 1.28, 2.29, 4.10, 7.32, 13.09, 23.40, 41.84, 74.79] to [0.43, 0.77, 1.37, 2.45, 4.38, 7.84, 14.01, 25.05, 44.78, 80.04] arcmin[13], the CFHTLenS angular values changed from [1.41, 2.79, 5.53, 11.0, 21.7, 43.0, 85.2] to [1.51, 3.00, 5.93, 11.77, 23.25, 45.77, 90.62] arcmin, and the KiDS-450 angular values changed from [0.71, 1.45, 2.96, 6.02, 12.25, 24.93, 50.75, 103.30, 210.27] to [0.77, 1.57, 3.19, 6.50, 13.25, 26.88, 54.50, 110.45, 219.36] arcmin. In Fig. A1 we show how this change in the angular bin values affect the cosmological constraints. In general, using the weighted bin centers give slightly higher S_8 values than using the bin centers.

APPENDIX B: THE KIDS-450 COVARIANCE

In Fig. B1 we show the effect of different approaches in marginalizing σ_m for the *Baseline* KiDS-450 case. For (1), we use the survey-provided covariance that includes σ_m. For (2) we use the survey-provided covariance without σ_m, and include σ_m using Eq. 12 of Hildebrandt et al. (2017). For (3), we use the survey-provided covariance without σ_m, and marginalize over σ_m at the parameter level. We find that (2) and (3) are in very good agreement, demonstrating that the two approaches are effectively equivalent, while (1) is shifted. The reason for the shift is due to the fact that (1) uses the noisy data vector in Eq. 12 of Hildebrandt et al. (2017), instead of a theoretical data vector. This is also explained in Troxel et al. (2018).
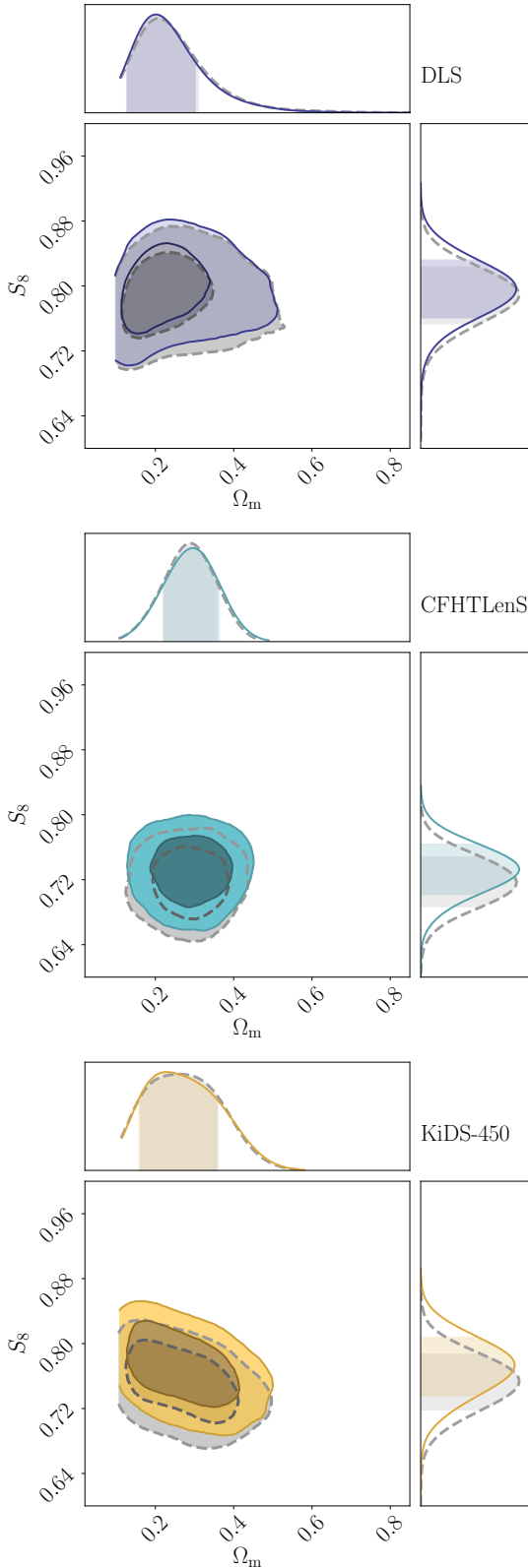
APPENDIX C: SCALE CUT VALUES

In Table C1 we list the scale cuts used in Sec. 5.3. We also show the redshift z_p used to calculate these scale cuts, as well as the number of data points remaining after making the scale cuts.
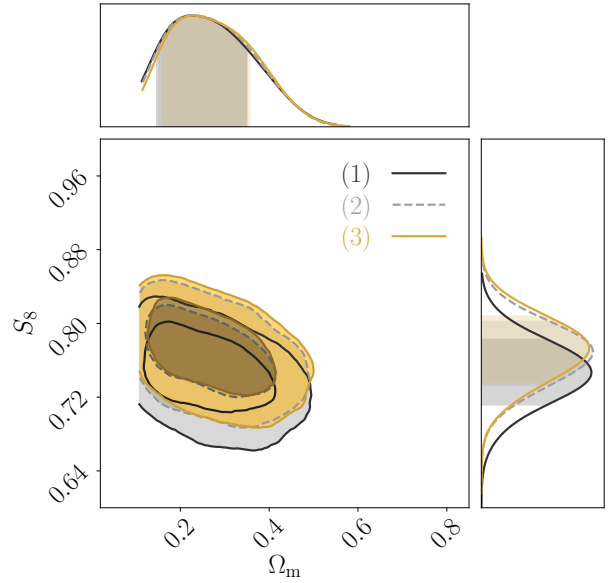
ACKNOWLEDGEMENT

[13] Since we do not have the catalogs, this was estimated using Eq. 9 of Krause et al. (2017).

Figure A1. This plot illustrates the change in the cosmological constraints when we change the angular values used in the data vector: the grey dashed contour shows the constraints using the bin centers, while the colored contour shows the constraints using the weighted centers.

Figure B1. KiDS-450 constraints when using different approaches to marginalize σ_m.

Below we acknowledge the data sources:

CFHTLenS: This work is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This research used the facilities of the

Table C1. *Top:* The redshift of the peak of the lensing efficiency for each redshift bin in each survey. z_p is used to calculate the scale cuts in the bottom table. *Bottom:* Scale cuts applied in each survey. Each row is for a combination of bins. The bins are ordered as (bin1, bin1), (bin1, bin2) ... (bin1, bin$N$), (bin2, bin2) ... (bin2, bin$N$) ... (bin$N$, bin$N$), where N is the maximum number of bins and bin$N$ is the highest redshift bin. The maximum scale cuts are fixed to the survey-specified scale cuts. "–" indicates no data points are used after the scale cut. The last row lists the remaining number of data points after the scale cut.

		DLS	CFHTLenS	DES-SV	KiDS-450
z_p	bin1	0.13	0.13	0.19	0.13
	bin2	0.22	0.18	0.29	0.23
	bin3	0.31	0.23	0.42	0.28
	bin4	0.42	0.28		0.38
	bin5	0.57	0.38		
	bin6		0.43		
	bin7		0.48		

	Conservative			
	DLS	CFHTLenS	DES-SV	KiDS-450
$\theta_{\mathrm{min},\pm}$ (arcmin)	8.7/78.6	8.7/78.6	6.2/56.5	9.0/81.3
	8.7/78.6	8.7/78.6	6.2/56.5	9.0/81.3
	8.7/78.6	8.7/78.6	6.2/56.5	9.0/81.3
	8.7/78.6	8.7/78.6	4.6/41.6	9.0/81.3
	8.7/78.6	8.7/78.6	4.6/41.6	5.6/50.4
	5.7/51.3	8.7/78.6	3.6/33.0	5.6/50.4
	5.7/51.3	8.7/78.6		5.6/50.4
	5.7/51.3	6.6/60.0		4.8/43.5
	5.7/51.3	6.6/60.0		4.8/43.5
	4.4/40.0	6.6/60.0		3.9/35.4
	4.4/40.0	6.6/60.0		
	4.4/40.0	6.6/60.0		
	3.6/33.0	6.6/60.0		
	3.6/33.0	5.5/49.6		
	3.1/28.0	5.5/49.6		
		5.5/49.6		
		5.5/49.6		
		5.5/49.6		
		4.8/43.0		
		4.8/43.0		
		4.8/43.0		
		4.8/43.0		
		3.9/35.1		
		3.9/35.1		
		3.9/35.1		
		3.6/32.6		
		3.6/32.6		
		3.4/30.6		
# of data points	94	137	25	52

The contributions from the primary authors are listed below. C.C. led the main analysis and writing of this paper. M.W. wrote the main software package WLPipe which integrates the python scripts using the Pegasus workflow engine. S.D. helped with the pipeline development, covariance assessment, comparison metrics, pipeline testing, and editing.