THE CATALOG-TO-COSMOLOGY FRAMEWORK FOR WEAK LENSING AND GALAXY CLUSTERING FOR LSST

J. Prat[*, 1, 2], J. Zuntz[†, 3], Y. Omori[1, 2], C. Chang[1, 2], T. Tröster[4], E. Pedersen[5], C. García-García[6],
E. Phillips-Longley[7], J. Sanchez[8], D. Alonso[6], X. Fang[9], E. Gawiser[10], K. Heitmann[11], M. Ishak[12], M. Jarvis[13],
E. Kovacs[11], P. Larsen[11], Y.-Y. Mao[14], L. Medina Varela[12], M. Paterno[15], S. D. P. Vitenti[16], Z. Zhang[1] and The
LSST Dark Energy Science Collaboration. *Author affiliations may be found before the references.*
Version April 25, 2023

ABSTRACT

We present TXPipe, a modular, automated and reproducible pipeline for ingesting catalog data and performing all the calculations required to obtain quality-assured two-point measurements of lensing and clustering, and their covariances, with the metadata necessary for parameter estimation. The pipeline is developed within the Rubin Observatory Legacy Survey of Space and Time (LSST) Dark Energy Science Collaboration (DESC), and designed for cosmology analyses using LSST data. In this paper, we present the pipeline for the so-called "3×2pt" analysis – a combination of three two-point functions that measure the auto- and cross-correlation between galaxy density and shapes. We perform the analysis both in real and harmonic space using TXPipe and other LSST-DESC tools. We validate the pipeline using Gaussian simulations and show that it accurately measures data vectors and recovers the input cosmology to the accuracy level required for the first year of LSST data under this simplified scenario. We also apply the pipeline to a realistic mock galaxy sample extracted from the CosmoDC2 simulation suite (Korytov et al. 2019). TXPipe establishes a baseline framework that can be built upon as the LSST survey proceeds. Furthermore, the pipeline is designed to be easily extended to science probes beyond the 3×2pt analysis.

Subject headings: methods: statistical – dark energy – large-scale structure of the universe

1. INTRODUCTION

The large-scale structure (LSS) contains rich information on both the geometry of spacetime and the growth of cosmic structure. Among the most direct avenues for probing the LSS is examining the statistical properties of the large-scale distribution of galaxies (Eisenstein et al. 2005; Springel et al. 2006), which are biased tracers of the distribution of mass. In addition, one can use the phenomenon of *weak gravitational lensing* – the small deflection of photon trajectories due to the perturbation of spacetime from mass – to map the distribution of mass directly. The weak lensing-inferred mass distribution is typically measured using the distortion of observed galaxy shapes (for a review of weak lensing, see e.g. Bartelmann & Schneider 2001). Recent analyses of galaxy surveys have further shown that it is even more effective to combine galaxy clustering and weak lensing in a multi-probe approach to jointly infer cosmology (e.g. DES Collaboration 2022b; Heymans et al. 2021).

In particular, a common approach is to combine three two-point functions of the galaxy density field δ_g and the weak lensing shear field γ: galaxy clustering $\langle\delta_g\delta_g\rangle$, galaxy-galaxy lensing $\langle\delta_g\gamma\rangle$ and cosmic shear $\langle\gamma\gamma\rangle$. In these measurements, we usually refer to the galaxy sample used for δ_g as the *lens galaxies*, and the sample used for weak lensing as the *source galaxies*. These two-point statistics capture the Gaussian information in the matter field, which is sensitive to cosmological parameters that describe the history and content of the Universe. The main advantage of combining the three probes in a coherent analysis is that since each probe depends on

cosmological and nuisance parameters in a different way, combining them allows us to effectively break the degeneracies between the parameters and tighten the overall cosmological constraints, as well as constraints on nuisance parameters. Following the community's convention, we refer to this combination of three two-point correlation functions as the 3×2pt probe.

Building on the success of the Stage-III[1] galaxy surveys the Dark Energy Survey (DES, Flaugher 2005), the Kilo-Degree Survey (KiDS, de Jong et al. 2013) and the Hyper Suprime-Cam survey (HSC, Aihara et al. 2018), we are at the very beginning of Stage-IV galaxy surveys with the Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić et al. 2019), the ESA satellite Euclid (Laureijs et al. 2011) and the NASA's Nancy Grace Roman Space Telescope (Akeson et al. 2019) ramping up their activities, and DESI (Levi et al. 2019) already operating. This paper in particular focuses on LSST, which is scheduled to begin its 10 year survey in 2024. The 3×2pt analysis is one of the baseline pillars of the LSST cosmology analysis and is expected to deliver a Dark Energy Task Force Figure of Merit (DETF FoM) ~ 60 (Albrecht et al. 2006b). Together with supernovae, galaxy clusters and strong lensing, LSST is expected to deliver a DETF FoM ~ 500 (The LSST Dark Energy Science Collaboration 2018).

To achieve these projections, one of the key factors will be our ability to control for systematic effects. The significantly lower statistical uncertainties in LSST com-

---

*: Corresponding author: jprat@uchicago.edu
†: Corresponding author: joe.zuntz@ed.ac.uk

[1] The Stage-III and Stage-IV classification was introduced in the Dark Energy Task Force report (Albrecht et al. 2006a), where Stage-III refers to the ongoing dark energy experiments that started in the 2010s and Stage-IV refers to those that start in the 2020s.

pared to current data places stringent requirements on the level of systematic effects that we can tolerate. At this point, the community collectively has extensive experience from Stage-III surveys where systematic effects that were previously ignored are becoming relevant, and this will only be more evident in LSST. Some examples for these "new" systematic effects include: the coupling of shear calibration bias and redshift uncertainties through the blending of object images (MacCrann et al. 2022); biases in photometric redshift distributions from incompleteness of spectroscopic training samples (Hartley et al. 2020); selection effects in galaxy samples used for clustering measurements (Pandey et al. 2022), to name only a few.

To this end, we cannot necessarily predict the next new systematic effect that will emerge, but we *can* learn from our Stage-III experiences and make the upcoming LSST cosmology analysis process smoother by building a framework to 1) minimize the human errors in carrying out the analysis and 2) enable the users to easily find and test systematic effects that could emerge. This is the main idea behind the creation of TXPipe[2]. In TXPipe we focus on the *measurement* side of the 3×2pt cosmology analysis, but note that its general structure could be applied to other analyses as well. We also note that similar concepts exist from the theory side in a number of common cosmology tools such as CosmoSIS (Zuntz et al. 2015) but to our knowledge it has not been developed at a similar level on the measurement side, which is equally important. In an earlier work in Chang et al. (2019), we explored a prototype measurement pipeline WLPipe[3] and applied it on four precursor datasets. The work demonstrated the value of such a framework by exposing a number of issues in earlier measurements. In a similar spirit, Longley et al. (2023) recently reanalyzed the DES Y1, HSC-Y1 and KiDS-1000 cosmic shear analyses using TXPipe, finding a few additional issues.

The goal of this paper is to first present the structure and design of TXPipe and then validate the basic functionalities using mock galaxy catalogs. Here we target the pipeline requirement for the 3×2pt cosmology analysis using the first year (Y1) of LSST data as it is the first near-term goal we expect from LSST. To perform the validation we use two sets of complementary mock galaxy catalogs: first, a set of simple Gaussian simulations to test that the basic measurements of the estimators do indeed recover the input signal; second, the DESC CosmoDC2 catalog presented and validated in Korytov et al. (2019) and Kovacs et al. (2022) to test the various functionalities in TXPipe that involve more realistic galaxy properties. We compare the measurements and the theory prediction both at the data vector level and at the level of the cosmological constraints. Whenever possible, we also validate the 3×2pt-related components of other DESC software packages with this exercise – these include FIRECrown[4] (likelihood), TJP-Cov[5] (covariance), the Core Cosmology Library (CCL[6], Chisari et al. 2019) and others. All together we are able

to demonstrate that the performance of core measurement components is sufficient for an LSST Y1 3×2pt cosmology analysis. We fully expect that the baseline analysis that will be adopted in later years of LSST will evolve depending on what we find, and so have designed TXPipe to be adaptable to future changes. We keep the constituent parts of the pipeline only loosely coupled to each other, so that changes can be isolated and their impacts carefully assessed. We store all the pipeline's outputs in one place, so that they can be compared one-to-one. And we maintain pipeline configurations under version control, and propagate this information to output metadata, so that the nature of changes made is never lost or unclear.

The paper is organized as follows. In Section 2 we introduce the basic background formalism used in a 3×2pt cosmology analysis. This includes both the theory prediction and the estimators used in this work. In Section 3 we present the design and functionalities of TXPipe. In Section 4 we validate TXPipe at both the data vector and the cosmology level using the LSST Y1 like Gaussian simulations and in Section 5 we apply it to CosmoDC2. We discuss lessons learned and future steps in Section 6 and summarize in Section 7.

2. MODELING: 3×2PT COSMOLOGY

The 3×2pt probes are the autocorrelation of the positions of galaxies, the cross-correlation of galaxy shapes and galaxy positions and the autocorrelation of galaxy shapes respectively. These measurements can be done both in real space and in harmonic space. We use two samples to perform these measurements, a *lens* sample for which we have position information and a *source* for which we have both position and shape information. We also use (true) redshift information for both samples. In this section we will describe how we define and model these two-point correlation functions.

Unless explicitly stated, during this study we use a cosmology close to the best-fit cosmological parameters from WMAP-7 (Komatsu et al. 2011): $\Omega_c = 0.22, \Omega_b = 0.048, h = 0.71, n_s = 0.963, \sigma_8 = 0.8$, which is the assumed cosmology in the CosmoDC2 simulation.

2.1. *Background theory*

Considering a real signal $a(\mathbf{n})$ on the unit sphere S^2, it is equivalently described in terms of its spherical harmonic coefficients $\{a_{\ell m}\}$ with $\ell \in \mathbb{N}$ and $m \in \{-\ell, ..., \ell\}$. The angular power spectrum $\{C_\ell^{ab}\}$ between two signals a and b is defined as:

$$\langle a_{\ell m} b_{\ell' m'}^* \rangle \equiv C_\ell^{ab} \, \delta_{\ell \ell'} \, \delta_{m m'}, \qquad (1)$$

and can be obtained using the following estimator assuming coverage over the full-sky:

$$C_\ell = \sum_{m=-\ell}^{\ell} \frac{|a_{\ell m}|^2}{2\ell + 1}. \qquad (2)$$

The angular power spectrum corresponds to the average power in fluctuations on scales of the order of π/ℓ on the sphere. If we assume that a is a random field, the power spectrum can be interpreted as a compression technique, and used to perform statistical inference on physical models of the field. In particular, if a is an

isotropic Gaussian random field, the power spectrum is a sufficient statistic containing all the relevant information in the realization a. In our case, a will be the galaxy number overdensity for galaxy clustering, and the convergence (or shear) field for weak lensing observables.

2.2. Harmonic and real space model

Under the Limber approximation (Limber 1953; LoVerde & Afshordi 2008) and assuming a flat Universe cosmology, we obtain the weak lensing *shear power spectrum* as a projection of the 3D matter power spectrum P_{mm}:

$$C_{\gamma\gamma}^{ij}(\ell) = \int d\chi \frac{q_s^i(\chi) q_s^j(\chi)}{\chi^2} P_{mm}\left(k = \frac{\ell+1/2}{\chi}, z(\chi)\right), \quad (3)$$

between two source redshift bins i, j where $q_s(\chi)$ are the window functions of the given source populations of galaxies, k is the 3D wavenumber, ℓ is the 2D multipole moment and χ is the comoving distance. We can also model the cross-correlation between the lens and source samples $C_{\delta_g\gamma}$ and express it as a projection of the 3D galaxy-matter power spectrum P_{gm}. For a lens redshift bin i with a window function $N_l(\chi)$ and a source redshift bin j,

$$C_{\delta_g\gamma}^{ij}(\ell) = \int d\chi \frac{N_l^i(\chi) q_s^j(\chi)}{\chi^2} P_{gm}\left(k = \frac{\ell+1/2}{\chi}, z(\chi)\right). \quad (4)$$

Finally the galaxy clustering harmonic space correlation function is a projection of the 3D galaxy power spectrum P_{gg}:

$$C_{\delta_g\delta_g}^{ij}(\ell) = \int d\chi \frac{N_l^i(\chi) N_l^j(\chi)}{\chi^2} P_{gg}\left(k = \frac{\ell+1/2}{\chi}, z(\chi)\right). \quad (5)$$

Limber's approximation holds if the 3D galaxy overdensity field of the lenses and the 3D matter overdensity field at the redshift of the source galaxies vary on length scales much smaller than the typical length scale of their respective window functions in the line of sight direction. The lens window function is defined as:

$$N_l^i(\chi) = \frac{n_l^i(z)}{\bar{n}_l^i} \frac{dz}{d\chi}, \quad (6)$$

where n_l^i is the lens redshift distribution and $\bar{n}_l^i$ is the mean number density of the lens galaxies. The weak lensing window function of the source galaxies is:

$$q_s^j(\chi) = \frac{3H_0^2\Omega_m}{2c^2} p(\ell) \frac{\chi}{a(\chi)} g(\chi), \quad (7)$$

where a is the scale factor, $p(\ell)$ is the ℓ-dependent prefactor in the lensing observables due to the spin-2 nature[7] of the shear and $g(\chi)$ is the lensing efficiency kernel:

$$g(\chi) = \int_\chi^{\chi_h} d\chi' \frac{n_s^j(z)}{\bar{n}_s^j} \frac{dz}{d\chi'} \frac{\chi'-\chi}{\chi'}, \quad (8)$$

with $n_s^j(z)$ being the redshift distribution of the source galaxies, $\bar{n}_s^j$ the mean number density of the source galaxies and χ_h is the comoving distance to the horizon.

---

[7] Here we use $p(\ell) = \ell^2/(\ell+1/2)^2$, which corresponds to the 1st order extended Limber flat projection (ExtL1Fl), as defined in table 1 from Kilbinger et al. (2017).

We use CCL (Chisari et al. 2019) to obtain the model for the two-point correlation functions. We compute the non-linear matter power spectrum using the Takahashi et al. (2012) version of HALOFIT. For the linear power spectrum, we use the CAMB algorithm (Lewis & Bridle 2002).

2.2.1. Galaxy bias model

In our fiducial model we assume that lens galaxies trace the mass distribution following a simple linear biasing model ($\delta_g = b\,\delta_m$, where b is modeled as a constant for each redshift bin), so the galaxy power spectrum and the galaxy-matter power spectrum relate to the matter power spectrum by different factors of the galaxy bias:

$$P_{gg} = b^2 P_{mm}, \quad (9)$$
$$P_{gm} = b\, P_{mm}. \quad (10)$$

2.2.2. Intrinsic alignment model

Intrinsic alignments (IAs) are due to correlations between the intrinsic galaxy shapes. IAs contribute to the total observed cosmic shear angular power spectra $C_{GG}^{ij}(\ell)$ with unknown additive terms of the form:

$$C_{GG}^{ij}(\ell) = C_{\gamma\gamma}^{ij}(\ell) + C_{II}^{ij}(\ell) + C_{\gamma I}^{ij}(\ell) + C_{\gamma I}^{ji}(\ell), \quad (11)$$

and to the total galaxy-shear angular spectra $C_{\delta_g G}^{ij}(\ell)$ with:

$$C_{\delta_g G}^{ij}(\ell) = C_{\delta_g\gamma}^{ij}(\ell) + C_{\delta_g I}^{ij}(\ell), \quad (12)$$

where G represents the total observed shape of the galaxies that include the cosmological shear γ and the intrinsic shape I. In Eqs. (3) and (4) we have defined the projections with respect to the true shear involving the matter power spectrum, here we define the rest of projections:

$$C_{II}^{ij}(\ell) = \int d\chi \frac{N_s^i(\chi) N_s^j(\chi)}{\chi^2} P_{II}\left(k = \frac{\ell+1/2}{\chi}, z(\chi)\right) \quad (13)$$

$$C_{\gamma I}^{ij}(\ell) = \int d\chi \frac{q_s^i(\chi) N_s^j(\chi)}{\chi^2} P_{\gamma I}\left(k = \frac{\ell+1/2}{\chi}, z(\chi)\right) \quad (14)$$

$$C_{\delta_g I}^{ij}(\ell) = \int d\chi \frac{N_l^i(\chi) N_s^j(\chi)}{\chi^2} P_{\gamma I}\left(k = \frac{\ell+1/2}{\chi}, z(\chi)\right) \quad (15)$$

where we have introduced the source window function $N_s^i(\chi)$ which is analogous to the lens one:

$$N_s^i(\chi) = \frac{n_s^i(z)}{\bar{n}_s^i} \frac{dz}{d\chi}. \quad (16)$$

The P_{II} and the $P_{\gamma I}$ power spectra are generic. Usually they are assumed to be linearly related to the local tidal field and to be of the same shape as the matter power spectrum except from a redshift dependent scaling:

$$P_{II}(k, z) = A(z)^2 P_{mm}(k, z), \quad (17)$$

$$P_{\gamma I}(k, z) = A(z) P_{mm}(k, z). \quad (18)$$

In this work we choose to use the two-parameter nonlinear alignment (NLA) model (Bridle & King 2007), which defines the amplitude parameter as:

$$A(z) = -A_{IA}\bar{C}_1 \frac{3H_0^2\Omega_m}{8\pi G} D^{-1}\left(\frac{1+z}{1+z_0}\right)^{\eta_{IA}}. \quad (19)$$

A_{IA} is one of the two free parameters of the NLA model and it is a dimensionless amplitude that governs the strength of the IA contamination. Here G is the gravitational constant and $D(z)$ is the linear growth factor. The normalization constant $\bar{C}_1$ is typically fixed at a value obtained from the SuperCOSMOS Sky Survey (Brown et al. 2002) of $\bar{C}_1 = 5 \times 10^{-14} M_\odot^{-1} h^{-2} Mpc^3$. η_{IA} is the other free parameter which controls the redshift scaling. z_0 is the pivot redshift which we fix to 0.62, which is a common choice.

2.2.3. Real space projection

Finally, each real-space two-point correlation function is related to the total angular power spectrum for cosmic shear, galaxy-galaxy lensing and galaxy clustering via

$$w^{i,j}(\theta) = \int \frac{d\ell\ell}{2\pi} C^{i,j}_{\delta_g \delta_g} J_0(\ell\theta), \tag{20}$$

$$\gamma_t^{i,j}(\theta) = \int \frac{d\ell\ell}{2\pi} C^{i,j}_{\delta_g G} J_2(\ell\theta), \tag{21}$$

and

$$\xi_{+/-}^{i,j}(\theta) = \int \frac{d\ell\ell}{2\pi} C^{i,j}_{GG} J_{0/4}(\ell\theta) \tag{22}$$

under the flat-sky approximation, where the J_α represent the Bessel functions of the first kind. We have tested that the flat-sky approximation is good enough for $w(\theta)$ given the LSST Y1 footprint and $\xi_{+/-}$ and γ_t are expected to be much less impacted by this effect (Kilbinger et al. 2017).

2.2.4. Redshift and shear marginalization

When obtaining cosmological constraints we marginalize over some observational systematics such as redshift and shear calibration effects to obtain more realistic posterior uncertainties. In particular, we marginalize over a shift in the mean redshift Δz^i both for the lens and source input redshift distributions n^i_{input}:

$$n^i(z) = n^i_{\text{input}}(z - \Delta z^i). \tag{23}$$

For the shear, we marginalize over a multiplicative shear bias m per each source bin, which modifies the shear and galaxy-shear angular power spectra in the following way:

$$C^{ij}_{\delta_g G}(\ell) = (1 + m^j) C^{ij}_{\delta_g G, \text{input}}(\ell) \tag{24}$$

$$C^{ij}_{GG}(\ell) = (1 + m^i)(1 + m^j) C^{ij}_{GG, \text{input}}(\ell) \tag{25}$$

Note that in this first validation of the DESC software we do not consider some effects found significant in recent data analyses such as magnification (Elvin-Poole & Mac-Crann et al., 2022) or redshift space distortions (Krause et al. 2021) (although these are now implemented in CCL).

3. TXPIPE

TXPIPE is DESC's implementation of the pipeline that ingests catalog data and performs all the calculations required to obtain quality-assured two-point measurements of lensing and clustering, and their covariances, with the metadata necessary for parameter estimation. The code is designed to collect and formalize the many calculations

and analysis stages that in previous surveys have often been manually connected.

The goal of the project is that the complete pipeline, from the output of the LSST Science Pipelines[8] to the inputs to cosmological parameter estimation, can be run and re-run in a single operation. Such unification is convenient, but it also permits provenance tracking of the pipeline to be essentially complete, so that the collaboration can be sure of precisely what code was run to generate data products.

We use continuous integration features to test the pipeline as changes are made, using *Github Actions*[9]. This system automatically runs a unit test suite whenever changes are proposed or made, and also a set of pipelines on 1 deg² of data. This is not large enough to ensure that the pipeline is numerically correct, but does ensure that core functions work.

In Figure 1 we show a flowchart that illustrates the TXPIPE pipeline in its most basic functionality. This flowchart is automatically generated by TXPIPE. The Figure shows the stages that have been used for the Gaussian simulation analysis which we present in Sec. 4. Besides the most basic stages presented and tested in this initial paper, in Section 3.1 we summarize additional features which are implemented within TXPIPE that will be needed for the data analysis.

Much of TXPIPE functionality involves connecting together external codes that perform measurements. The package wraps each of these tools as a Python class, specifying input and output files required for each, and launching them using the CECI[10] library and executable which automatically interfaces them to workflow management frameworks. TXPIPE also performs various calculations internally, again using a Python class to embody each pipeline stage. The outputs of the pipeline can be automatically organized as a web page. It is also important to highlight that in all these cases TXPIPE is optimized for the large data volumes of LSST. Wherever possible we use parallel and online algorithms, which do not require all data to be loaded in memory at once and are therefore scalable. All correlation function outputs are saved using the SACC[11] format, a dedicated DESC software for storing measurements, covariances and redshift distributions in a unified way. Below we describe the TXPIPE pipeline stages grouped in different blocks:

- **Data ingestion, shear and redshift calibration:** The TXPIPE pipeline starts with the ingestion of the galaxy catalogs which are prepared into suitable formats (e.g. HDF5 files for catalogs). TXPIPE then splits the catalogs into different tomographic bins, typically using criteria described below in Section 5. Next, it builds the redshift distributions[12], performs the requested shear calibration using the methods of METADETECTION and METACALIBRATION described in Sheldon et al. (2020), Sheldon &

---

[8] https://pipelines.lsst.io/
[9] https://docs.github.com/en/actions
[10] https://github.com/LSSTDESC/Ceci
[11] https://github.com/LSSTDESC/sacc
[12] In this work we use true redshift distributions throughout, making this step a simple histogram. However, implementations are undergoing to make this step realistic. See the additional functionalities section in 3.1 for more details.
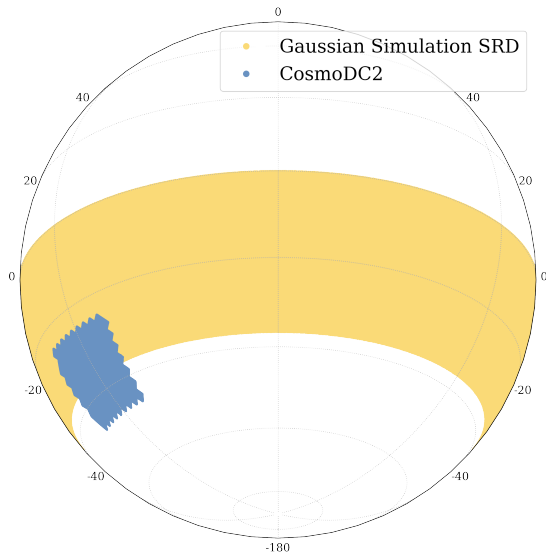
FIG. 1.— TXPipe flowchart for the Gaussian simulation pipeline. The inputs are represented by yellow boxes and are typically HDF5 or YAML files, the TXPIPE stages by red ellipses and each of the outputs by blue boxes, which are typically HDF5, SACC, PNG or PDF files. The input and output formats can be defined by the user.

Huff (2017) and LENSFIT described in Miller et al. (2007), depending on the catalog type.

- **Two-point measurements:** We use the harmonic-space and real-space estimators described in Appendix A. In essence, for the harmonic-space estimator, we first create shear and density maps in HEALPix[13] formats. Then we use the NAMASTER[14] (Alonso et al. 2019) pseudo-C_ℓ code to measure the spectra. For the real-space estimators, we invoke the fast tree code TREECORR[15] (Jarvis et al. 2004) to perform the measurements. A mask is needed for the harmonic space estimator, which we generate directly using a depth map obtained from the lens galaxies (using a resolution much coarser than the average separation between galaxies). This is a simplistic approach that will need to be updated once we test on more complex simulations or data. We also generate a random catalog from the depth maps to be used in the real-space estimator (see Equations A13 and A14). Typically, the number of random points is set to be at least 20 times larger

than that of the lens galaxies (Prat et al. 2022). This puts a significant memory load on the real-space measurements given the number of lenses expected for LSST Y1. As a result, we also implement a pixel-based estimator described in Sec. A.2.1 – our validation tests for real-space galaxy clustering and galaxy-galaxy lensing measurements were performed using this pixel-based estimator.

- **Two-point predictions:** TXPIPE interfaces with CCL to obtain a theory prediction for the corresponding measured two-point correlation functions, automatically using the same angular binning as the measurements, i.e. generating a *theory data vector* that we use to compare with the measurements. The same theory predictions are also used in the covariance matrix calculation described below.

- **Covariance matrix:** TXPIPE obtains a covariance matrix for the corresponding data vector using one of several codes with a unified interface in the DESC TJPCov package. Currently TJPCov includes Gaussian covariances with different treatments of the mask. We compare the different options in Section 4.2 and Appendix B.

[13] https://healpix.sourceforge.io/
[14] https://github.com/LSSTDESC/NaMaster
[15] https://github.com/rmjarvis/TreeCorr

FIG. 2.— Sky coverage or masks of the two simulations we use in this work: a Gaussian simulation mimicking a general LSST-Y1 like mask with 12,300 deg^2 and CosmoDC2 with 440 deg^2.

FIG. 3.— Redshift distributions of the tomographic bins for the lens (top) and source (bottom) samples as specified in the LSST-DESC SRD for the first year of LSST operation and for the samples we have defined on the CosmoDC2 simulation. The shaded regions represent the bin edges for the CosmoDC2 samples, which are designed to 1) match the SRD binning for the lenses and 2) produce a similar number density in each source bin.

- **Extra diagnostics:** TXPipe generates useful diagnostic plots and metadata throughout the pipeline, including plots of maps, catalog histograms, masks, jackknife patches, $n(z)$ distributions as well as quantities such as galaxy number densities, shape noise values, mask area etc[16].

[16] The full list of currently implemented diagnostics

3.1. Additional functionalities

Most of the validation tests performed in this paper assume an idealized LSST Y1 dataset. We also test TX-Pipe on the more realistic mock catalog CosmoDC2 in Section 5, but still with idealized conditions. Thus, we only test the core functionalities of TXPipe. There are additional functionalities that have been developed for more realistic analyses and that will continue to be developed and tested in future work. We briefly describe them here:

- **Redshift distribution estimation:** TXPipe is integrated with the DESC photometric redshift code RAIL tool[17], which contains a suite of algorithms to estimate both point-estimate redshifts per galaxy[18] and redshift distribution estimates for ensembles of galaxies[19]. The former are sometimes used to define tomographic bins (although, see Zuntz et al. 2021, for alternative binning approaches), while the latter are used for making theory predictions for cosmological inference.

- **Null tests:** In most Stage-III surveys, the measurement of the data vectors are accompanied by a large suite of null tests to ensure that the data vector is not significantly contaminated by systematic effects (see, for example Gatti et al. 2021; Giblin et al. 2021; Li et al. 2022; Rodríguez-Monroy et al. 2022a). These tests include correlations of shear and density with survey property maps and catalogs, testing for B-mode leakage, calculating mean shear in bins of size and signal-to-noise, etc. In the idealized simulations used in this paper, we do not put in any systematic effects, thus these functionalities will not be rigorously tested here and we leave this for future work.

- **Blinding:** In the era of precision cosmology, it is important to build in some mechanism so that we do not base our analysis choices on the outcome of our measurements. These so-called "observer biases" can be minimized by "blinding" the analysis and only "unblinding" the results once all analysis choices are frozen. In TXPipe, we currently implement the methodology described in Muir et al. (2020), where the data vectors are shifted to a slightly different cosmology in a way that preserves the relation between all parts of the data vector.

- **Convergence maps:** TXPipe interfaces with the WLMassMap tools[20], where the weak lensing shear catalogs are converted to convergence maps using a range of methods (e.g. Jeffrey et al. 2021). These maps are typically used for higher-order statistics such as peak counts (Liu et al. 2015;

can be found in https://txpipe.readthedocs.io/en/latest/stages/Diagnostics.html
[17] https://github.com/LSSTDESC/RAIL
[18] Including BPZ, Delight, FlexZBoost, and PZFlow (Benítez 2000; Coe et al. 2006; Leistedt & Hogg 2017; Izbicki & Lee 2017; Dalmasso et al. 2020; Crenshaw & Doster 2022)
[19] Including self-organizing maps, direct calibration, and variational-inference based models (e.g. Carrasco Kind & Brunner 2014; Wright et al. 2020; Rau et al. 2021)
[20] https://github.com/LSSTDESC/WLMassMap

Zürcher et al. 2022) and moments (Gatti et al. 2020, 2022). This functionality highlights the flexibility of TXPipe, where data vectors beyond 3×2pt can be incorporated into the framework and share the same infrastructure.

In general, it is straightforward to build on the existing TXPipe structure and incorporate new functionalities thanks to its modular and transparent design. We expect TXPipe to grow and become more complete/mature as the LSST data arrives in the coming years.

3.2. *Performance requirements*

Because of its depth and area, LSST data volumes will be very large: the final 10-year data release will contain $O(10^{10})$ galaxies. This size, and the corresponding increases in required systematic accuracy, mean that many methods and algorithms must be re-designed for LSST. DESC's primary computing facility is the National Energy Research Scientific Computing Center (NERSC), and we have targeted the parallelization features to deal with this size at the systems they host.

Our primary limitation is memory; many algorithms begin by loading entire catalog columns before performing operations on them, or on subsets on them. At full LSST scale a single column can be more than 50 GB in size. Although loading these is possible on high-memory systems, depending on the number of columns needed, doing so limits the number of processes that can simultaneously operate on data. We minimize such patterns in TXPipe wherever possible, preferring to model our data tables as streams and processing them in parallel. Since many operations involve core descriptive statistics on data columns (means, standard deviations, etc), we use a DESC library[21] of tools to implement them where possible.

Given the large data set size, processing speed is also greatly important. In many cases I/O is the bottleneck in TXPipe stages. The complete analysis described here iterates through the complete input catalogs multiple times; some algorithmic steps in the pipeline inherently require multi-pass runs, although sometimes it is a choice. We make heavy use of NERSC's LUSTRE file system's parallel I/O facilities, which *stripe* chunks of data files across multiple storage targets, and enable fast access to different parts of the data from different processes in parallel. Storing data in the HDF5 format makes it straightforward to both read and write different data chunks in parallel through MPI, provided that data chunks are large and contiguous.

We use a hybrid of Message Passing Interface (MPI) process and OpenMP (thread) parallelism paradigms. For example, when assigning objects to tomographic bins in `TXSourceSelector`, a low-CPU and thus I/O-dominated operation, we use pure MPI, creating as many processes as there are CPUs on a node, and splitting data among them. In cases where I/O is not the bottleneck, or where we can pre-reduce data down to multiple smaller subsets, such as when computing correlation functions in `TXTwoPoint`, we typically make use of thread parallelism

with OpenMP, creating only a single process per node with many threads.

A step further in this case is the use of GPU acceleration with TXPipe on for example the NERSC environment. GPU acceleration can speed up numerical calculations that involve a large throughput. We tested here an implementation of the TXNoiseMaps using the Google Jax library stage of TXPipe so that it can benefit from the speed up allowed by GPU acceleration offered by the NERSC Perlmutter system. The TXNoiseMaps stage generates random rotations and applies them to galaxy shape catalogs. By default, the stage evaluates rotations of galaxy shears, 100,000 items at a time. We find that for large enough catalogs, the speedup from using GPU acceleration is significant (of the order of a factor of 10).

There is sometimes a trade-off to be made between wider algorithmic speed and code flexibility and modularity. We can try to minimize the total number of I/O passes of the data, at the cost of requiring much greater coordination and unified behavior between stages. We avoid this where stages are not conceptually connected in TXPipe, which also enables us to remain flexible regarding stage input and output data. For example, we have a pipeline stage that does diagnostic plots on the source sample (such as shear as a function of object size, and similar null tests), and another stage that assigns objects to tomographic bins. Both use the same input data - the shear catalog - so could be combined in a single step. We choose not to do so, however, so that we can modify and verify them independently. However, we do not go to the extreme length of splitting the diagnostics stage into multiple stages, each for different null tests, since the small gain in clarity would not be useful enough to offset the cost of the time taken for added loops through the data.

We specify the computation time and resources we use for each TXPipe stage when running on the LSST-Y1 like Gaussian simulation in Table 1. Note that this catalog actually has a much larger effective number density for the lens sample due to reasons specified in Sec. 4.1. Therefore, the times detailed in this table are expected to be an upper limit for the data runs.

4. VALIDATION OF TXPIPE

In this section we validate the core functionalities of TXPipe to show that it meets the accuracy/precision requirement needed for the first year of LSST data as specified by the LSST DESC Science Requirement Document (hereafter the DESC SRD, The LSST Dark Energy Science Collaboration 2018). First we introduce the mock galaxy catalog used for the validation (Section 4.1). Next we describe our covariance matrix estimation (Section 4.2). We then validate the measurement of the data vectors (Section 4.3) and the recovered cosmological constraints (Section 4.4). All validations are done for the 3×2pt probes in both harmonic and real space.

4.1. *Gaussian mock galaxy catalog*

We generate a set of idealized Gaussian simulations following the galaxy sample specifications described in the LSST DESC SRD to stress-test TXPipe. Such a setup with perfectly known inputs allows us to validate whether the pipeline is capable of recovering our input

TXPIPE Stage	Time	N_n	N_p	N_t	Description
TXConvergenceMaps	38 min	1	1	32	Make a convergence map from a source map the using Kaiser-Squires method.
TXDensityMaps	1 min	1	1	1	Convert galaxy count maps to overdensity maps.
TXFourierGaussianCovariance	2 min	1	1	64	Compute a Gaussian Fourier-space covariance with TJPCov using f_sky only.
TXGaussianSimsMock	49 min	1	1	20	Simulate mock photometry from Gaussian simulations.
TXJackknifeCenters	8 min	1	1	1	Generate jackknife centers from random catalogs.
TXExternalLensMaps	132 min	2	10	1	Make tomographic lens number count maps from external lens catalog.
TXRealGaussianCovariance	19 min	1	1	64	Compute a Gaussian real-space covariance with TJPCov using f_sky only.
TXShearCalibration	7 min	1	7	1	Split the shear catalog into calibrated bins.
TXSourceMaps	51 min	4	4	10	Make tomographic shear maps from shear catalogs and tomography.
TXSourceSelectorMetadetect	11 min	2	64	1	Source selection and tomography for metadetect catalogs.
TXTwoPointFourier	32 min	4	4	20	Make Fourier space 3×2pt measurements using NaMaster.
TXTwoPointPixel	780 min	6	6	32	Compute pixelated versions of the 3×2pt real space correlation functions.
TXTwoPointTheoryFourier	1 min	1	1	1	Compute theory predictions for Fourier-space 3×2pt measurements.
TXTwoPointTheoryReal	0.2 min	1	1	1	Compute theory predictions for real-space 3×2pt measurements.
TXUniformDepthMap	0.1 min	1	1	1	Generate a uniform depth map from the mask.

TABLE 1
BRIEF DESCRIPTION OF EACH STAGE FOR GAUSSIAN SIMULATIONS PIPELINE AND TIME IT TAKES TO RUN. N_n IS THE NUMBER OF NODES, N_p IS THE TOTAL NUMBER OF PROCESSES AND N_t THE NUMBER OF THREADS PER PROCESS. SEE HTTPS://TXPIPE.READTHEDOCS.IO/EN/LATEST/STAGES.HTML FOR THE LATEST DOCUMENTATION OF EACH STAGE.

Lens bin	$\langle z \rangle$	Number density	Galaxy bias
1	0.30	2.25	1.229
2	0.50	3.11	1.362
3	0.70	3.09	1.502
4	0.90	2.61	1.648
5	1.10	2.00	1.799

Source bin	$\langle z \rangle$	Number density	Shape noise σ_e
1	0.31		
2	0.49		
3	0.69	1.78	0.26
4	0.96		
5	1.59		

TABLE 2
LSST Y1 DESC SRD SAMPLE SPECIFICATIONS. THE NUMBER DENSITIES ARE IN $1/\mathrm{ARCMIN}^2$ AND σ_e IS DEFINED IN EQ. 28. THE SAME NUMBER DENSITY AND SHAPE NOISE IS ASSUMED IN ALL SOURCE REDSHIFT BINS.

signal to the precision required for LSST-Y1 analyses. Once this baseline is validated, we can then move on to more realistic mock galaxy catalogs from e.g. N-body simulations (see Section 5). We show the sky coverage for each simulation in Fig. 2 using the CARTOSKY[22] code.

To generate the mock catalog, we use the redshift distributions shown in Figure 3 with five lens redshift bins and five source redshift bins, which follow the definitions given in the DESC SRD. We also assume the number densities listed in that document, which are 10 gal/arcmin^2 for the source catalog and 18 gal/arcmin^2 for the lens catalog, but note that these are the total number densities if one were to integrate over the full redshift distributions ignoring the tomographic bins. We use the redshift distributions to obtain the binned number densities which are listed in Table 2.

Assuming the cosmology from Section 2, we generate theory predictions of all the 3×2pt auto and cross C_ℓ data vectors in harmonic space using CCL up to $\ell_\mathrm{max} = 16384$. We use the galaxy bias values listed in Table 2. From the C_ℓ's we generate correlated, noiseless Gaussian maps at a HEALPix resolution $N_\mathrm{side} = 8192$[23] for both spin-0

fields (density, δ_m) and spin-2 fields (shear, γ) using the approach described in Giannantonio et al. (2008). To turn the noiseless maps into galaxy catalogs, we employ the following process:

- We apply a simple mask to define the survey footprint, illustrated in Fig. 2. We apply a declination cut of $-36.61° < \mathrm{Dec} < 0°$, which results in an area of 12,300 deg^2, consistent with the DESC SRD specification for the LSST Y1 data for the 3×2pt analysis. Note that the actual LSST footprint will cover a larger range of declinations but maintain the same area due to regions of high Galactic dust as detailed in Lochner et al. (2022, 2018).

- To generate source galaxies, we randomly sample points inside the survey mask with a number density according to the DESC SRD as listed in Table 2. For each galaxy, we obtain a true shear value $\gamma = \gamma_1 + i\gamma_2$ corresponding to the pixel in the shear maps it falls in. Then, this catalog is ingested into TXPIPE, which adds shape noise to it. To do that, we draw from a Gaussian distribution with a width $\sigma = 0.26$ to add a random "intrinsic shape" component. We draw independent values for $\gamma_{\mathrm{int},1}$, $\gamma_{\mathrm{int},2}$, which represent the intrinsic shape of each galaxy $\gamma_\mathrm{int} = \gamma_{\mathrm{int},1} + i\gamma_{\mathrm{int},2}$. We then add γ_int to the true shear γ to form the observed ellipticity values e (Seitz & Schneider 1997):

$$e = \frac{\gamma_\mathrm{int} + \gamma}{1 + \gamma^* \gamma_\mathrm{int}}, \qquad (26)$$

from which we extract each ellipticity component as $e_1 = \Re[e]$, $e_2 = \Im[e]$.

- To generate lens galaxies, we Poisson sample the density map with the galaxy number density listed in Table 2. Since a density field cannot have values smaller than -1, the Gaussian field will have

harmonic space and real-space mimicking the process we would apply to real data, we needed to use a high ℓ_max in order for the real space input theory to also be accurate. We found that in this case there is a significant aliasing of power to lower ℓ, which produces a significant mismatch of the resulting measurements with the input theory, unless the resolution is high enough.

[22] https://github.com/kadrlica/cartosky
[23] Note that even though later we only use large scales in the analysis we still need a high N_side value because of the following reasoning: Since the simulations were designed to work both for

tails that cannot be Poisson sampled. To circumvent this, we scale the density maps for each tomographic bin by factors 1/[2.458, 2.043, 1.878, 2.060, 2.249] such that the fraction of pixels with values < -1 is small ($< 1\%$), and verify that the imposed cut does not induce differences in the power spectra. Specifically, these numbers come from imposing that $b\delta_m > -1$ for each redshift bin. The galaxy bias b is larger for the higher redshift bins, while the matter fluctuations δ_m are larger for the lower redshift bins, thus compensating each other and yielding a similar scaling factor in all redshift bins. The field is then Poisson sampled with the desired galaxy number density to form the lens sample. Due to this scaling, the galaxy clustering and galaxy-galaxy lensing data vectors measured from this catalog will be artificially low and will need to be rescaled by the same factor to recover the correct amplitude (in the case of galaxy clustering by the square of these factors), as has been done before in e.g. Elvin-Poole et al. (2018). Note that the target number density (shown in Table. 2) needs to be multiplied by the scaling factor squared when sampling from the field, otherwise the resulting shot noise levels would be much higher due to this needed scaling of the data vectors. This results in approximately [603 M, 576M, 482 M, 490M, 449M] galaxies per each lens bin and 20 times more random points (for comparison the source catalog has $\simeq$ 78M objects per source bin).

- After the galaxy catalogs are generated, they are fed to TXPipe in a similar way as we would input real data catalogs. Then, we use TXPipe stages to generate the galaxy and shear maps that are used to compute the two-point correlation functions using the estimators detailed in Appendix. A. We use $N_{\rm side} = 4096$ for all the maps generated within TX-Pipe.

4.2. *Covariance matrix*

To evaluate the accuracy of TXPipe against theoretical predictions, and to do inference, we need an estimate of the covariance matrix. In this work we use a Gaussian covariance matrix. The Gaussian component coming from the 2-point functions is dominant with respect to the non-Gaussian parts related to using higher order information from the bispectrum. In harmonic space, at fixed cosmology, the Gaussian component is approximately described by (Schneider et al. 2002; Crocce et al. 2011)

$$\mathcal{C}[C_X^{ij}(\ell), C_{X'}^{i'j'}(\ell')] = \delta_{\ell\ell'} \frac{D_X^{ii'}(\ell)D_X^{jj'}(\ell) + D_X^{ij'}(\ell)D_X^{ji'}(\ell)}{(2\ell+1)f_{\rm sky}}, \quad (27)$$

where i, j and i', j' denotes the redshift bin pairs associated with the two considered power spectra; X is either $\gamma\gamma$, $\delta_g\gamma$ or $\delta_g\delta_g$. $D_X^{ij}(\ell) \equiv C_X^{ij}(\ell) + N_X^{ij}(\ell)$ is the sum of the signal C_X^{ij} (described in Section 2.2) and noise power spectra N_X^{ij}, $\delta_{\ell\ell'}$ is the Kronecker delta function and $f_{\rm sky}$ is the fractional sky coverage. The noise power spectra is $1/n_{\rm gal}$ for $C_{\delta_g\delta_g}$ and $\sigma_e^2/n_{\rm gal}$ for $C_{\gamma\gamma}$ and are assumed to be zero for cross-correlations between different redshift bins and

	Harmonic space	Real space
Lens bin		
1	$\ell < 253$	$\theta > 42.81$ arcmin
2	$\ell < 402$	$\theta > 26.93$ arcmin
3	$\ell < 535$	$\theta > 20.21$ arcmin
4	$\ell < 654$	$\theta > 16.52$ arcmin
5	$\ell < 761$	$\theta > 14.21$ arcmin
Source bin		
1–5	$\ell < 3000$	$\theta > 2.5$ arcmin

TABLE 3
SCALE CUTS IN REAL SPACE AND HARMONIC SPACE, THAT IN THE CASE OF THE LENS SAMPLE CORRESPOND TO $k_{\rm MAX} = 0.3h/\rm Mpc$, ACCORDINGLY WITH THE DESC SRD SPECIFICATIONS.

for the galaxy-shear cross-correlation. σ_e is the shape noise per component defined as

$$\sigma_e \equiv \sqrt{\frac{1}{2}\left(\frac{{\rm var}(e_1)}{R_{11}^2} + \frac{{\rm var}(e_2)}{R_{22}^2}\right)} \quad (28)$$

within the METACALIBRATION framework for a diagonal response matrix. R_{ii} is the response factor for the ellipticity component i. For the Gaussian simulations, we assume an identity response matrix and thus we recover $\sigma_e = 0.26$, the same value we input as the standard deviation of the ellipticity per component.

In practice, the covariance is non-Gaussian and cosmology-dependent. There are plans to implement these improvements for the analysis with future LSST data. However, for the analysis presented here these approximations are sufficient – see figure 3 from Friedrich et al. (2021) which shows that the non-Gaussian terms represent less than 10% of the diagonal elements of the covariance for both cosmic shear and angular galaxy clustering. The validation of these terms with DESC software is left for future work.

To convert Equation 27 into a real-space covariance we apply the Hankel transform operator to the Fourier space covariance, implementing the approach in Singh (2021) and SKYLENS[24].

In Equation 27 and also in the real space transformation we have made the assumption that the precise geometry of the footprint of the dataset does not have a big effect on the covariance, and we use the simple $f_{\rm sky}$ factor to account for the size of the footprint. As was shown in Troxel et al. (2018), this can lead to biases in the covariance when the footprint is small. However, we check in Appendix B that the effect is small for our setup for the LSST Y1-like Gaussian simulation. In particular we find the diagonal errorbars without including this effect are slightly smaller, thus making our validation tests if anything more stringent. For the CosmoDC2 simulation this effect becomes important, especially at large scales. Thus, for the tests in CosmoDC2 we include the mask effects in the harmonic space covariance using the DESC TJPCov package, see Appendix B and Section 5 for more details.

4.3. *Data vector*

In Figures 4, 5 and 6 we show the data vector outputs of TXPipe for cosmic shear, galaxy-galaxy lensing and galaxy clustering, in real and harmonic space. In

---

[24] https://github.com/sukhdeep2/Skylens_public
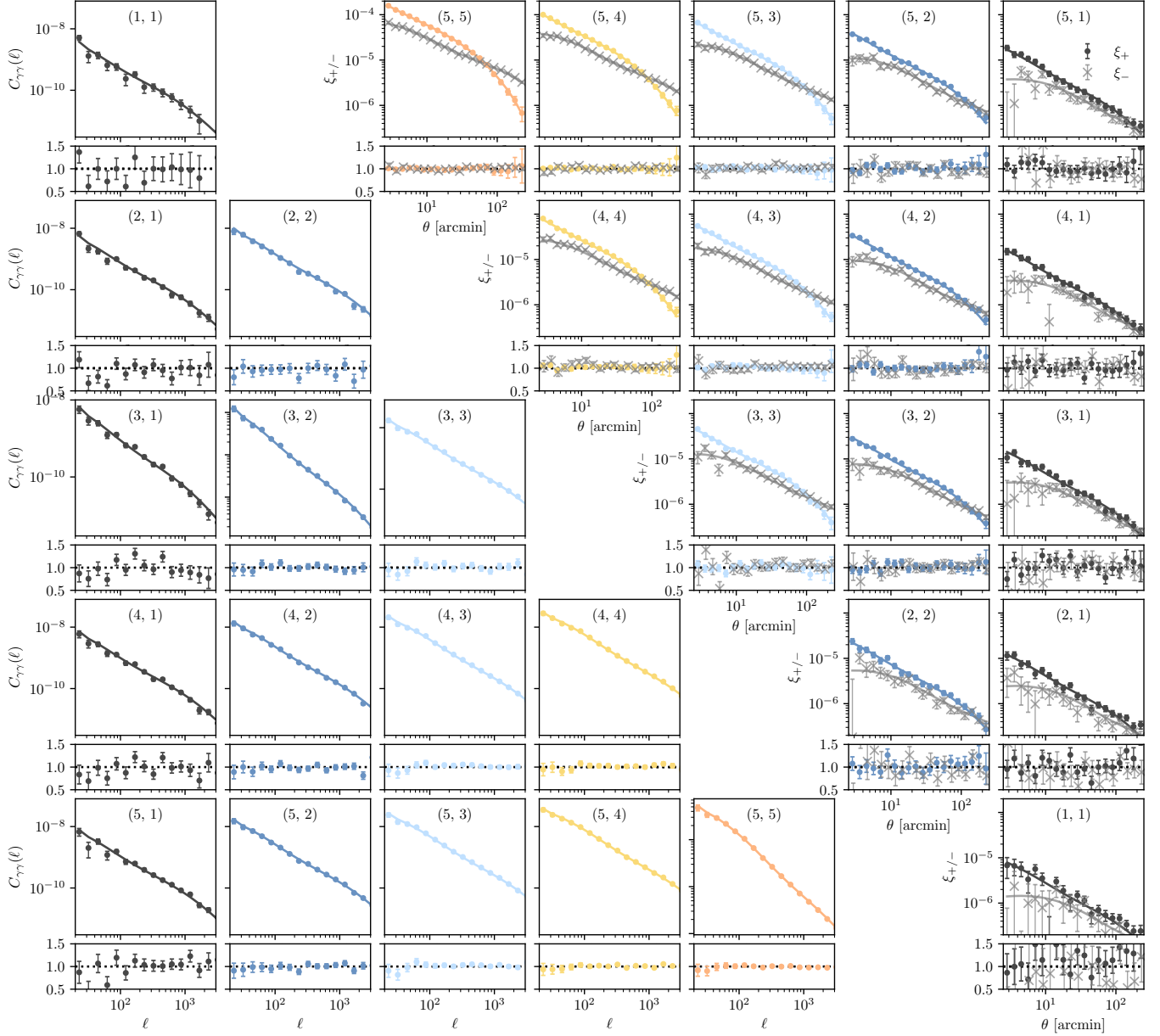
Prat, Zuntz et al. (LSST DESC)

FIG. 4.— Cosmic shear measurements (correlation between galaxy weak lensing and galaxy weak lensing) from TXPipe in harmonic space ($C_{\gamma\gamma}$, lower left) and real space (ξ_+, ξ_-, upper right). Each panel corresponds to a (source, source) redshift bin combination. For each combination, we show the measurements with point markers from our Gaussian simulations and the input theory with lines (upper panels) and the ratio of the measurement to the theory (lower panels). The error bars are the square root of the diagonal covariance.

	Probe	χ^2/ν (r1)	PTE (r1)	χ^2/ν (r2)	PTE (r2)
Real space	ξ_+	$259.9/300 = 0.87$	0.95	$260.8/300 = 0.87$	0.95
	ξ_-	$283.7/300 = 0.95$	0.74	$284.7/300 = 0.95$	0.73
	γ_t	$91.5/67 = 1.37$	0.02	$60.8/67 = 0.91$	0.69
	w	$57.8/53 = 1.09$	0.30	$71.3/53 = 1.35$	0.05
	3×2pt	$722.6/720 = 1.00$	0.47	$680.3/720 = 0.94$	0.85
Harmonic space	$C_{\gamma\gamma}$	$232.8/225 = 1.03$	0.35	$223.9/225 = 1.00$	0.51
	$C_{\delta_g\gamma}$	$75.3/63 = 1.20$	0.14	$56.3/63 = 0.89$	0.71
	$C_{\delta_g\delta_g}$	$55.5/49 = 1.13$	0.24	$40.2/49 = 0.82$	0.81
	3×2pt	$383.2/337 = 1.14$	0.04	$339.7/337 = 1.01$	0.45

TABLE 4
χ^2 PER DEGREE OF FREEDOM AND THE PROBABILITY-TO-EXCEED (PTE), AFTER SCALE CUTS, WHEN COMPARING TXPIPE OUTPUT AND INPUT FOR THE DESC SRD LSST Y1-LIKE GAUSSIAN SIMULATION. WE LIST THE RESULTS FOR TWO DIFFERENT REALIZATIONS: R1 AND R2. FOR THE REST OF THE PAPER WE ALWAYS USE THE DATAVECTORS FROM R1.
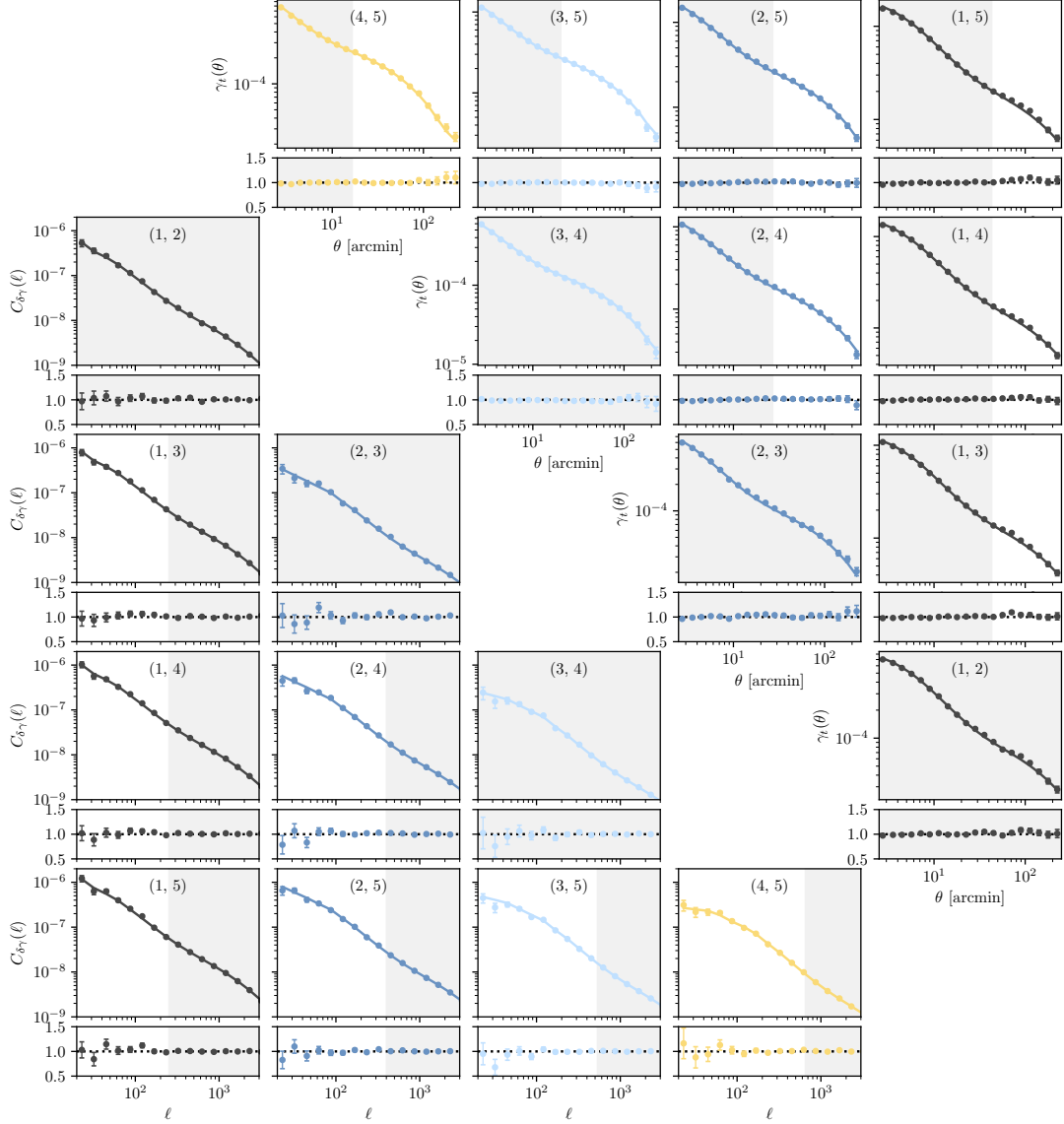
FIG. 5.— Galaxy-galaxy lensing measurements (correlation between galaxy position and galaxy weak lensing) from TXPIPE in harmonic space (lower left) and real space measurements (upper right). Each panel corresponds to a (lens, source) redshift bin combination. For each lens-source bin pair, we show the measurements from our Gaussian simulations and the input theory (upper panels) and the ratio of the measurement to the theory (lower panels). The error bars are the square root of the diagonal covariance. For simplicity, we do not show all the bin pairs that are measured, but only the highest signal-to-noise ones, which are the ones included in the cosmological analysis, following the DESC SRD. In addition, we show the SRD scale cuts represented with the gray shaded bands for scales we do not use – some of the bins are entirely removed.
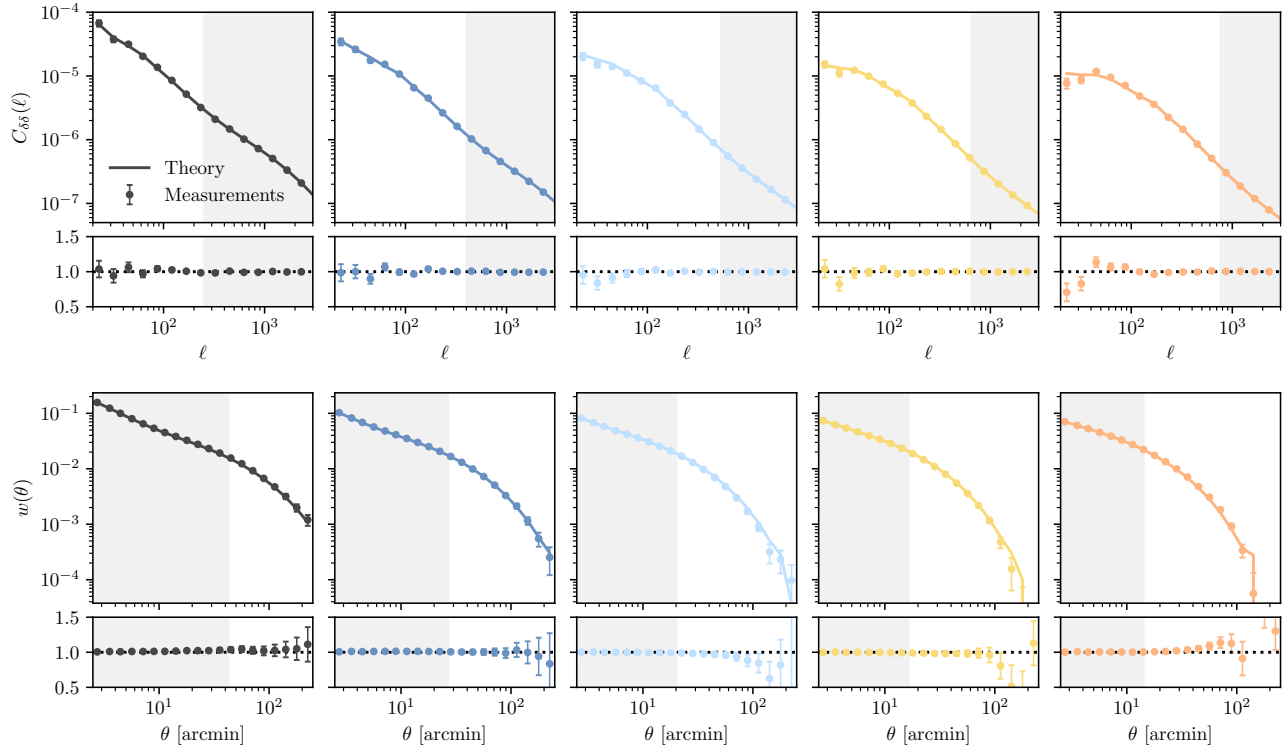
FIG. 6.— Galaxy clustering measurements from TXPIPE in harmonic (top row) and real space (bottom row). We only show the auto-correlation measurements, which are the ones included in the cosmological analyses, following the DESC SRD specifications. For each bin, we show the measurements from our Gaussian simulations and the input theory (upper panels) and the ratio of the measurement to the theory (lower panels). The error bars are the square root of the diagonal covariance. The gray shaded bands represent the SRD scale cuts.

all the figures we also show the input theoretical prediction as well as the ratio of our measurements to the theory. For cosmic shear we show all the bin combinations. For galaxy-galaxy lensing and galaxy clustering, we only show a subset of the bin combinations for simplicity, in particular the ones where the signal-to-noise is larger and that are included in our analysis. The error bars in these figures are the square root of the diagonal of the covariance matrix described in Section 4.2. Our measurements are performed on scales $2.5 < \theta < 250$ arcmin for real space (with 20 logarithmic bins) and $20 < \ell < 3722$ for harmonic space (with 17 logarithmic bins, to match the scales defined in the DESC SRD). In all ongoing galaxy surveys, the modeling of the 3×2pt data vector is uncertain on small angular scales due to either uncertainty in the matter power spectrum $P_{\rm mm}$ or the uncertainty in nonlinear galaxy bias, which affects both $P_{\rm gg}$ and $P_{\rm gm}$. These scale cuts are currently one of the determining factors of the constraining power of a given dataset. We adopt the DESC SRD scale cuts listed in Table 3, with $k_{\rm max} = 0.3\,h\,{\rm Mpc}^{-1} = 0.213\,{\rm Mpc}^{-1}$ for the lenses used in galaxy clustering and galaxy-galaxy lensing and $\ell < 3000$ for cosmic shear. We convert from the 3D quantity k to the projected ℓ using:

$$\ell_{\rm max} = k_{\rm max}\chi(z_l) - 0.5. \qquad (29)$$

For the galaxy-galaxy lensing and galaxy clustering probes, we convert to a real space scale cut using the approximation $\theta_{\rm min} \sim \pi/\ell_{\rm max}$. Moreover, for galaxy-galaxy lensing we only use the lens-source redshift bin combinations that are indicated in the DESC SRD, which only include: $z_l^1-z_s^3$, $z_l^1-z_s^4$, $z_l^1-z_s^5$, $z_l^2-z_s^4$, $z_l^2-z_s^5$, $z_l^3-z_s^5$ and $z_l^4-z_s^5$,

corresponding to the combinations with higher signal-to-noise and where the lenses and the sources barely overlap in redshift. For galaxy clustering we only include the auto-correlations, as specified in the SRD. Note the choice of redshift combinations and of all the scale cuts might need to be revised for the analysis with data. In particular: a) overlapping bins between the lenses and sources will probably need to be included in future analyses since they are useful to self-calibrate the intrinsic alignment parameters, as it was done in e.g. the DES Y1 and in the DES Y3 3×2pt analyses (Abbott et al. 2018; DES Collaboration 2022b), and b) clustering cross-correlations between redshift bins are useful to self-calibrate the redshift distributions of the lens sample (Nicola et al. 2020; Hang et al. 2021). For cosmic shear, we include all the scales we measure in real space and we cut at $\ell < 3000$ in harmonic space, following the SRD. In general, within the scale cuts, we find excellent agreement between the measurements and the input. To quantify this agreement, we evaluate the χ^2 per degree of freedom ν defined as

$$\chi^2/\nu = \frac{1}{\nu}(D_{\rm TXPipe} - D_{\rm theory})^t \mathcal{C}^{-1}(D_{\rm TXPipe} - D_{\rm theory}), \qquad (30)$$

where ν is the number of data points, $D_{\rm TXPipe}$ is the data vector as measured by TXPIPE, $D_{\rm theory}$ is the theory data vector evaluated at the same angular scales as the measurements[25] and $\mathcal{C}$ is the covariance matrix. We list the

[25] In harmonic space we weight the scales using the bandpower window functions within the NAMASTER formalism, as detailed in Appendix A. In real space we evaluate the theory at the mean angular scale value, as measured in the data using the *meanlogr*

Parameter	Prior
Cosmological parameters	
Ω_c	$\mathcal{N}[0.220, 0.2]$
Ω_b	$\mathcal{N}[0.0448, 0.006]$
σ_8	$\mathcal{N}[0.8, 0.14]$
$A_s \times 10^{-9}$	$\mathcal{N}[2.16, 0.378]$
h	$\mathcal{N}[0.71, 0.063]$
n_s	$\mathcal{N}[0.963, 0.08]$
w_0	$\mathcal{N}[-1, 0.8]$
w_a	$\mathcal{N}[0.0, 2.0]$
Astrophysical nuisance parameters	
$b^{1\cdots5}$	$\mathcal{N}[1.229, 0.9], \mathcal{N}[1.362, 0.9], \mathcal{N}[1.502, 0.9],$
	$\mathcal{N}[1.648, 0.9], \mathcal{N}[1.799, 0.9]$
A_{IA}	$\mathcal{N}[0.0, 3.9]$
η_{IA}	$\mathcal{N}[0.0, 2.3]$
Observational nuisance parameters	
$m^{1\cdots5}$	$\mathcal{N}[0.0, 0.013]$
$\Delta z_l^{1\cdots5} \times 10^3$	$\mathcal{N}[0.0, 6.6], \mathcal{N}[0.0, 7.5], \mathcal{N}[0.0, 8.5],$
	$\mathcal{N}[0.0, 9.5], \mathcal{N}[0.0, 10.5]$
$\Delta z_s^{1\cdots5} \times 10^3$	$\mathcal{N}[0.0, 2.6], \mathcal{N}[0.0, 3.0], \mathcal{N}[0.0, 3.4],$
	$\mathcal{N}[0.0, 3.9], \mathcal{N}[0.0, 5.2]$

TABLE 5
PRIORS FOR THE PARAMETERS OF OUR MODEL. $\mathcal{N}(\mu, \sigma)$ IS A GAUSSIAN DISTRIBUTION WITH A MEAN μ AND A WIDTH σ. THE PRIORS MOSTLY FOLLOW THOSE SPECIFIED IN THE LSST DESC SRD, BUT CENTERED ON THE TRUE VALUES OF OUR SIMULATIONS.

χ^2/ν values in Table 4. We also report the Probability-To-Exceed (PTE) or also sometimes called p-value, defined as:

$$\text{PTE} = 1 - \int_0^{\chi_D^2} \chi^2(x, \nu_D) dx, \qquad (31)$$

where the χ^2 function is integrated until the given χ_D^2 value for the data set D.

Note that we find some of the PTE values to be below 0.05 or above 0.95. This is expected given a distribution of values but we still want to check that they are indeed due to noise and not due to an identifiable issue. In order to do so we run a second realization of the Gaussian simulations, for which we also list the results in Table. 4. We find that most of the deviations that are below 0.05 or above 0.95 in the first realization yield a good agreement (values between 0.16 and 0.84) in the second realization, except for the PTE 0.95 value for ξ_+ that remains the same in both realizations, pointing to a possible overestimation of the covariance for this probe and also potentially for ξ_-. Besides this case, we conclude that the rest of the anomalies are due to statistical fluctuations.

4.4. Cosmology inference

We now take the data vectors measured from TXPIPE and propagate them into cosmological constraints to validate that we can recover the input cosmology to the required precision and accuracy. To this end, we use the DESC likelihood package FIRECROWN[26]. In this work we assume a Gaussian likelihood throughout. The core of FIRECROWN uses CCL as its theory prediction and it is designed to interface with established cosmological inference code such as COBAYA and COSMOSIS so that we can take advantage of the fast sampling and caching

`TreeCorr` function.

[26] https://github.com/LSSTDESC/firecrown

techniques implemented therein. This is the first time FIRECROWN is used for a scientific project, and is a suitable test case given the idealized settings. We use the EMCEE sampler (Foreman-Mackey et al. 2013), and obtain 640,0000 samples for each chain, for which we apply a burn-in of 150,000 samples in all cases. We use CHAIN-CONSUMER (Hinton 2016) with a Kernel Density Estimation (KDE) smoothing value of 2.0 for the plots and to obtain the mean and 1-σ values shown in the tables.

We fit the measured data vectors to a ΛCDM model and a $w_0 - w_a$CDM model with the priors on the cosmological parameters and nuisance parameters listed in Table 5, which closely follows that used in the DESC SRD – see Appendix C for a detailed description of the differences between our analysis and DESC SRD. In Fig. 7 we show the cosmological results from FIRECROWN for the harmonic space analysis using the data vectors derived from TXPIPE in Section 4.3. We show the cosmic shear only results, the galaxy-galaxy lensing and galaxy clustering (also called 2×2pt) results and their combination. In Figures 8 and 9 we also compare the constraints coming from the real and harmonic space estimators for a subset of the parameters, which we find to be in very good agreement. We show the rest of the parameters for these cases in Appendix D, together with the nuisance parameter posteriors. We also present a summary of the results in Table 6 for ΛCDM and in Table 7 for $w_0 - w_a$CDM. In all cases, we recover the true input cosmological parameters within the 1σ contours.

We use the APRIORI sampler provided by COSMOSIS within FIRECROWN to sample the prior and understand which parameters we are able to constrain with respect to the prior. We add the APRIORI samples in all the contour plots as the gray dotted lines. Overall we find that with LSST Y1 precision all the cosmological parameters are significantly constrained with respect to the prior in the 3×2pt combination. Also, as shown in Fig. 7 we find that cosmic shear gives more precise results for n_s and S_8, while the 2×2pt combination is able to constraint Ω_m and Ω_b more tightly.

Moreover, using the priors stated in Table 5 we find that both the ΛCDM and the $w_0 - w_a$CDM LSST-Y1 analyses are systematics limited, meaning that the constraints on S_8 improve significantly when fixing observational systematics. Specifically, we find the constraints on S_8 are $\sim$5 times more precise in ΛCDM and $\sim$1.8 more precise in $w_0 - w_a$CDM times when fixing the shear and redshift calibration parameters, while the parameters describing the equation of state of dark energy do not vary as much and yield similar constraints. In Fig. 16 we display this comparison.

Then, we also compare our results on the LSST Y1 patch simulated using the SRD specifications with the forecast on cosmological parameters from the DESC SRD. Generally we find the constraining power of our posteriors is similar to what was predicted in the SRD. This is an important cross-check given that the software tools used in the two studies are significantly different. For more details we refer the reader to Appendix C.

We also compare the results from this work with the most constraining Stage-III posteriors to date, the 3×2pt constraints from the first three years (Y3) of DES data (DES Collaboration 2022b,a). We find that the LSST Y1 results on the Gaussians simulation provide a constraint
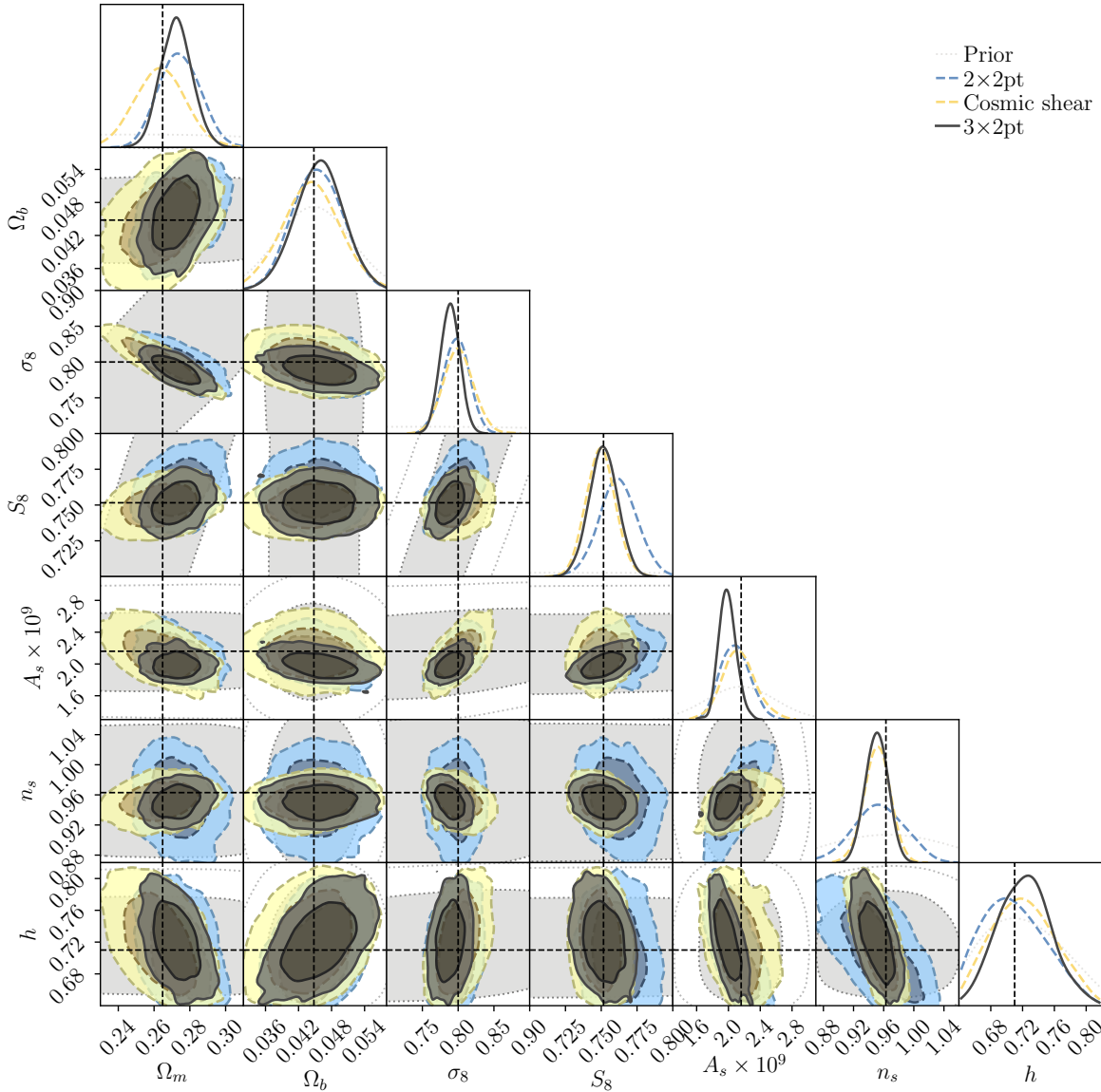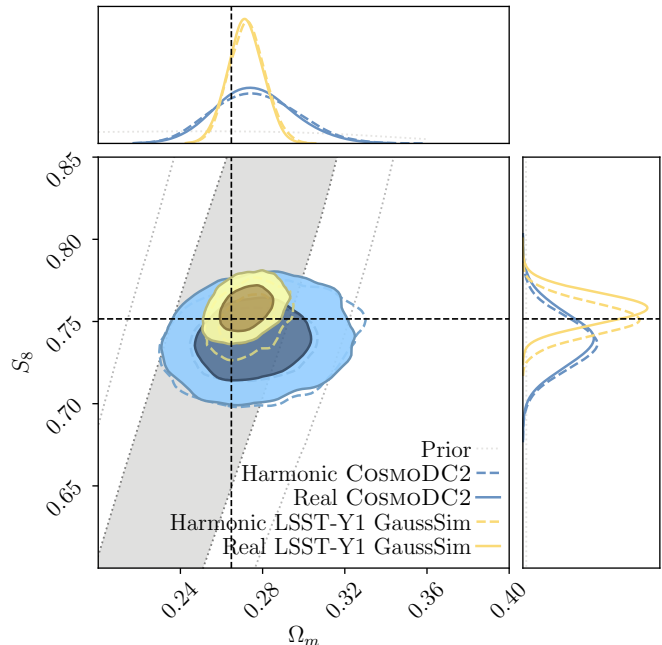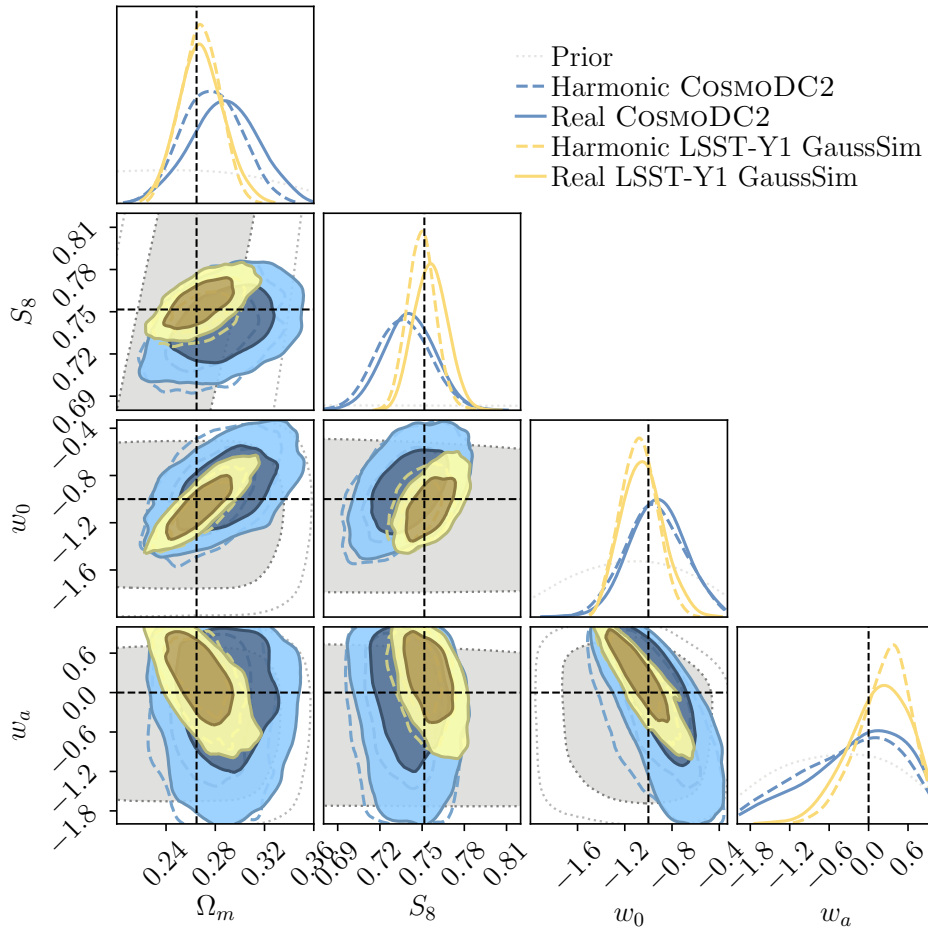
FIG. 7.— Cosmological contours assuming the ΛCDM model for the LSST Y1-like Gaussian mock catalogs compared with the input values and the DESC SRD priors. We show the results for the harmonic space 2×2pt analysis (galaxy clustering and galaxy-galaxy lensing, *dashed blue*), cosmic shear only (*dashed yellow*) and their combination 3×2pt (*solid black*).

in S_8 that is $\simeq 2$ times tighter than DES Y3 in ΛCDM and $\simeq 2.5$ times in $w_0 - w_a$CDM, as well as a significant improvement in the rest of the cosmological parameters. Note that this comparison is not straightforward because of two main reasons: 1) The priors from the DES Y3 analysis are uninformative (flat) while we use Gaussian priors. 2) The IA model is not the same for the DES Y3 ΛCDM and our work (we marginalize over a the 2-parameter NLA model and the DES Y3 ΛCDM analysis marginalizes over a 5-parameter TATT model). The IA models *are* the same in both analyses for $w_0 - w_a$CDM, which provides a cleaner comparison that we show in Figure 15. Therefore, even with these caveats in mind, we expect that the LSST Y1 data set will be able to shed light into the current S_8 tension, see e.g. Di Valentino

et al. (2021).

Overall, the work presented here shows that the DESC tools that are needed to perform an end-to-end 3×2pt analysis are sufficiently accurate in their most basic functionalities for LSST Y1 precision. This includes analyses both in harmonic space and in real space, from the data vector measurement to the parameter inference.

5. APPLICATION TO COSMODC2

Now that we have validated the core functionalities of TXPIPE at scale, we describe its application to a more complex mock galaxy catalog as a first step towards applying the pipeline to data (we note that in Longley et al. 2023, TXPIPE was used with Stage-III data, but that paper focused on testing cosmic shear in real space, while

ΛCDM	σ_8	Ω_m	S_8	FoM
Harmonic Gauss. Sims. LSST Y1	$0.789^{+0.013}_{-0.014}$	$0.2724^{+0.0086}_{-0.0092}$	$0.7511^{+0.0104}_{-0.0097}$	12,317
Harmonic Gauss. Sims. LSST Y1 fixed obs. sys.	$0.7920^{+0.0092}_{-0.0089}$	$0.2715^{+0.0062}_{-0.0067}$	$0.7532^{+0.0019}_{-0.0017}$	91,760
Real Gauss. Sims. LSST Y1	$0.797^{+0.012}_{-0.014}$	$0.2711^{+0.0091}_{-0.0081}$	$0.7582^{+0.0092}_{-0.0095}$	13,604
Real Gauss. Sims. LSST Y1 fixed obs. sys.	$0.7935^{+0.0079}_{-0.0091}$	$0.2721^{+0.0062}_{-0.0058}$	$0.7554^{+0.0020}_{-0.0019}$	89,928
Harmonic CosmoDC2	0.767 ± 0.031	-	0.737 ± 0.016	-
Real CosmoDC2	$0.771^{+0.031}_{-0.029}$	-	$0.740^{+0.016}_{-0.018}$	-

TABLE 6
RESULTS FOR THE LSST Y1 GAUSSIAN SIMULATION AND COSMODC2 UNDER THE ΛCDM MODEL, SHOWN ALSO IN FIG. 8. WE ALSO ADD THE RESULTS FIXING THE OBSERVATIONAL SYSTEMATICS PARAMETERS, WHICH INCLUDE MULTIPLICATIVE SHEAR BIASES AND PHOTOMETRIC REDSHIFTS (SEE APPENDIX C). THE FIGURE OF MERIT IS DEFINED AS $\mathrm{FoM} = \det[\mathcal{C}(\Omega_m, \sigma_8)]^{-1/2}$, WHERE $\mathcal{C}$ IS THE COVARIANCE OF THE POSTERIOR PARAMETERS.

wCDM	σ_8	Ω_m	S_8	w_0	w_a	FoM
Harm. Gauss. Sims. LSST Y1	$0.793^{+0.020}_{-0.022}$	$0.267^{+0.017}_{-0.015}$	$0.7500^{+0.0093}_{-0.0108}$	$-1.08^{+0.15}_{-0.17}$	$0.37^{+0.32}_{-0.47}$	33.1
Harm. Gauss. Sims. LSST Y1 fixed obs. sys.	$0.804^{+0.012}_{-0.014}$	$0.259^{+0.011}_{-0.010}$	0.7470 ± 0.0057	$-1.20^{+0.14}_{-0.12}$	$0.66^{+0.28}_{-0.48}$	49.4
Real Gauss. Sims. LSST Y1	$0.800^{+0.022}_{-0.023}$	$0.267^{+0.019}_{-0.018}$	$0.756^{+0.012}_{-0.013}$	$-1.05^{+0.17}_{-0.19}$	$0.23^{+0.50}_{-0.46}$	27.2
Real Gauss. Sims. LSST Y1 fixed obs. sys.	$0.801^{+0.014}_{-0.013}$	0.265 ± 0.012	$0.7531^{+0.0067}_{-0.0061}$	$-1.14^{+0.17}_{-0.12}$	$0.59^{+0.31}_{-0.54}$	42.0
Harmonic CosmoDC2	$0.766^{+0.033}_{-0.031}$	-	0.736 ± 0.020	-	-	-
Real CosmoDC2	$0.753^{+0.038}_{-0.031}$	-	$0.740^{+0.019}_{-0.017}$	-	-	-
Forecasted in DESC SRD LSST Y1	0.831 ± 0.020	0.315 ± 0.020	0.852 ± 0.010	-1.00 ± 0.21	0.00 ± 0.69	17.5

TABLE 7
ANALOGOUS RESULTS FOR THE $w_0 - w_a$CDM MODEL, SHOWN ALSO IN FIG. 9. IN THE LAST ROW WE ALSO ADD THE FORECAST PRODUCED BY THE DESC SRD ASSUMING THE SAME AREA AS OUR GAUSSIAN SIMULATION. THE CENTER VALUES OF THAT FORECAST WERE DIFFERENT BUT THE UNCERTAINTIES CAN BE COMPARED. THE FIGURE OF MERIT IS DEFINED AS $\mathrm{FoM} = \det[\mathcal{C}(w_0, w_a)]^{-1/2}$, WHERE $\mathcal{C}$ IS THE COVARIANCE OF THE POSTERIOR PARAMETERS.

Lens bin	$\langle z \rangle$	Number density	Galaxy bias
1	0.30	2.48	0.87
2	0.50	3.51	1.02
3	0.70	4.11	1.19
4	0.90	4.12	1.30
5	1.10	2.49	1.54
Source bin	$\langle z \rangle$	Number density	Shape noise σ_e
1	0.37	2.85	0.288
2	0.52	2.87	0.317
3	0.66	3.35	0.305
4	0.83	5.00	0.335
5	1.29	7.79	0.353

TABLE 8
COSMODC2 SAMPLE SPECIFICATION. THE NUMBER DENSITIES ARE IN $1/\mathrm{ARCMIN}^2$ AND σ_e IS DEFINED IN EQ. 28. WE ALSO LIST THE GALAXY BIAS VALUES THAT WE INPUT TO THE COVARIANCE. WE CHECK THEY ARE CONSISTENT WITH THE POSTERIORS IN FIGURE 10.

we test the full 3×2pt data vector in both real and harmonic space here). In particular, we run TXPIPE on the LSST DESC simulation suite CosmoDC2 (Korytov et al. 2019; Kovacs et al. 2022).

CosmoDC2 is a mock galaxy catalog based on the Outer Rim N-body simulation (Heitmann et al. 2019). It covers ~ 440 deg^2 area to redshift and depth much beyond that expected for LSST. Each galaxy is assigned observable quantities from photometry to morphology based on a number of semi-analytical models. The galaxies also carry cosmological information from both their position on the sky and ray-traced weak lensing properties. This provides a fairly realistic test bed for TXPIPE while still having a known true input cosmology to the simulations.

We add noise to the CosmoDC2 observed magnitudes following the methodology described in Ivezic et al. (2010), which accounts for the width of the point spread function (PSF), sky brightness, instrumental noise, and

FIG. 8.— Cosmological contours assuming the ΛCDM model for the CosmoDC2 and LSST Y1-like Gaussian mock catalogs compared with the input values and the DESC SRD priors that we use in both cases.

pixelization[27]. We use 16 visits per band per pixel, approximating an LSST Y1 survey. We also include the noise and estimator response on the shear truth values, using a constant shape noise per component i of $\sigma = 0.26$ and shear estimator response values (Sheldon et al. 2020).

[27] Future noise simulations will be done with updated values of the parameters in that document, which may be found at https://smtn-002.lsst.io/

FIG. 9.— Cosmological contours assuming the w_0-w_aCDM model for the CosmoDC2 and LSST Y1-like Gaussian mock catalogs compared with the input values and the DESC SRD priors that we use in both cases.

We sample the responses from a multivariate normal with the mean and standard deviation from a spline in size and signal-to-noise fitted to the DES-Year 1 catalog (Zuntz et al. 2018), and a correlation coefficient $\rho = 0.2$ between the size and shape response, as measured from the DES-Year 1 shape catalog. We do not vary any noise properties across the field. Using Eq. 28, we obtain the shape noise values per component σ_e displayed in Table 8.

We construct a similar lens sample as that used in Section 4.1 based on the DESC SRD, using all galaxies with $i < 24.1$. For source galaxies we select objects on the *riz* signal-to-noise and size as measured by the trace of the moments matrix compared to that of the PSF: $\text{SNR} > 10, T/T_{\text{psf}} > 0.5$ (see Zuntz et al. 2018 for more details). For both samples we use a random forest algorithm to assign galaxies to tomographic redshift bins, training a classifier on an ideal spectroscopic sample. For the lens sample we use *grizy* for the training while for the source sample the training can only be done in the *riz* bands because of the requirements of metacalibration - for more details on both the algorithm and this limitation see Zuntz et al. (2021).

The resulting redshift distribution and galaxy characteristics are shown in Figure 3 and Table 8. We note that this is not identical to the specification in DESC SRD, while it is consistent with the findings in Zhang

et al. (2022b). We next run TXPIPE with this sample. CosmoDC2 offers us additional opportunities to test TXPIPE as well as the simulations. For instance, we are able to test effects such as non-unity METADE-TECTION responses (see Section A.3), non-linear effects in the matter power spectrum on small scales, generating tomographic bins based on the galaxy colors, and many more.

Then, we estimate the linear galaxy bias of the lens sample performing a direct fit to the large scales for galaxy-galaxy lensing and galaxy clustering. We report these values in Table 8. Using these galaxy bias values, we obtain a good agreement between the measurements and the theory data vectors both in real and harmonic space after applying the scale cuts detailed in Table 3. We display the resulting χ^2 values in Table 9. We also show the two-point measurements in Appendix E.

We use these bias values to infer the Gaussian covariance matrix and use FIRECROWN to perform parameter inference. Then, we check the galaxy bias posteriors are consistent with the input values, which we demonstrate in Figure 10. Since they are found consistent, we do not update our initial estimates in the covariance. Note that the difference between the galaxy bias we find in Cos-MODC2 and what we input for the Gaussian sims might have an impact on the relative constraining power of each
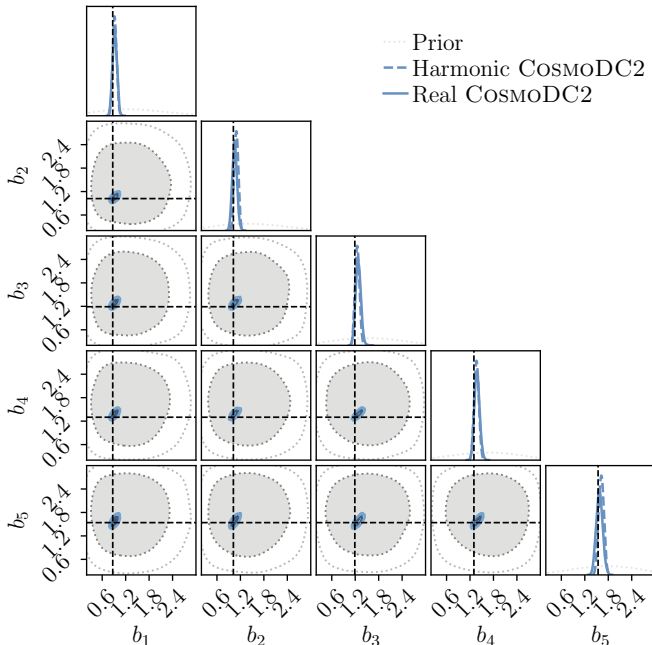
FIG. 10.— Linear galaxy bias for the CosmoDC2 lens catalog. The dashed lines mark the values we have used as input to the covariance matrix and for the χ^2 values.

	Probe	χ^2/ν	PTE
	ξ_+	$295.7/300 = 0.99$	0.56
	ξ_-	$297.5/300 = 0.99$	0.53
Real space	γ_t	$67.4/67 = 1.01$	0.46
	w	$48.8/53 = 0.92$	0.64
	3×2pt	$716.8/720 = 1.00$	0.53
	$C_{\gamma\gamma}$	$200.4/225 = 0.89$	0.88
Harmonic space	$C_{\delta_g\gamma}$	$47.9/63 = 0.76$	0.92
	$C_{\delta_g\delta_g}$	$45.3/49 = 0.92$	0.62
	3×2pt	$296.4/337 = 0.88$	0.95

TABLE 9
χ^2 VALUES WHEN COMPARING TXPIPE OUTPUT AND INPUT FOR COSMODC2. FOR COSMODC2 INPUT WE USE THE GALAXY BIAS VALUES LISTED IN TABLE 8 AND THE INPUT COSMOLOGY TO THE SIMULATIONS.

simulation. However, since the area difference is so large, we expect this will not be a dominant effect.

The final cosmological constraints are shown in Figure 8 for the ΛCDM model and in Figure 9 for $w_0 - w_a$CDM. Comparing with the priors, we find we are only able to constraint σ_8 and S_8, which we list in Tables 6 and 7 – the rest of the parameters are prior-dominated, as shown in Figure 15. In the latter, we also compare them with the DES Y3 3×2pt results assuming the same $w_0 - w_a$CDM model.

6. DISCUSSION: LESSONS LEARNED AND FUTURE WORK

Lessons learned— Several practical lessons can be learnt from this validation exercise relevant for LSST-scale datasets, some of which would not have been obvious in Stage-III datasets:

- Monitoring intermediate steps is critical for multiple parts of the pipeline, and devising a mechanism to do this automatically requires careful choices.

In many cases, such as when validating maps generated in our pipeline, a key test is the general question "does anything look unusual in this map?" rather than a purely numerical comparison that is not easily automated. We also have a number of null tests that we can run automatically on data, to ensure that, for example, galaxy shear is not correlated with signal-to-noise. But we have enough of these tests that we do expect some to randomly fail at the 2σ, even on correct data, so we cannot automate this process entirely.

- The complex web of external software dependencies in umbrella frameworks like TXPipe makes testing challenging; continuous integration is a relatively small part of addressing this problem, and more complete regression tests on large data sets were needed. Changes in code or packaging of other DESC projects like RAIL and general packages like NUMPY have required careful version management. We found that a containerization approach using Docker (Merkel 2014) is effective for management, using two images, one for experimentation and one more stable, and both built automatically using Github Actions.

- Having redundancy in the pipeline allows for very powerful cross-checks and debugging and we should incorporate this as much as possible into our pipelines. This includes multiple estimators, multiple implementations of covariance and noise estimations, etc. For example, as shown in Fig. 12 and detailed in Appendix A5 we have implemented two ways of estimating the noise power spectrum needed for the Fourier space measurements. This comparison revealed some subtle issues related to handling maps with weights per object and a partial mask.

- Similarly, comparing different parts of the TXPipe output to external codes is an essential exercise that often reveals problems that would otherwise be hard to track. Even though this kind of tests are especially needed for validating new code, periodic comparisons can expose unexpected behaviors from different revisions of the code. In this work, we carried out several comparisons of different outputs such as: (i) Real space two-point measurements and theory data vectors, comparing them with another DESC code DESCQA (see Appendix F for examples of the aftermath of such a comparison); (ii) harmonic space two-point measurements and theory data vectors, comparing them with a personal code (iii) real space analytical covariance with the external code CosmoCov[28] (Krause & Eifler 2017; Fang et al. 2020) and with the TXPipe's jackknife covariance (iv) harmonic space analytical covariance with a covariance obtained from several gaussian simulation realizations and with an external personal code from one of the authors. All these comparisons exposed issues in each of the pieces and helped to disentangle errors in such a large

---
[28] https://github.com/CosmoLike/CosmoCov

pipeline when the PTE validation tests were not meeting our criteria.

- It is extremely valuable to validate pipelines on simple simulations such as the Gaussian simulation used in this work before embarking on tests with more realistic N-body simulations, in order to be able to disentangle between different issues in an easier way. Even this nominally simple set-up has allowed us to find multiple issues and inaccurate approximations. In Appendix F we list some examples of this, to illustrate such a process.

- Once the pipeline has been tested in the simplest setup it becomes equally essential to test it with a more realistic scenario that mimics the data better. For example, non-unity lens and source galaxy weights and non-unity response functions are prone to introduce many bugs if not thoroughly tested.

- The usage of real-space randoms to account for mask effects becomes unsustainable at LSST scale, given the memory load it needs. Instead, pixel-based estimators are much more efficient in this regime. More generally, the right memory vs speed trade-offs may change significantly at LSST scale vs. what has been found in Stage III galaxy surveys.

Future work— We have introduced a number of simplifications in this work in order to carry out clean and unambiguous tests of the basic pipeline. Moving forward, we will need to improve on several areas described below in order to bring the DESC infrastructure to the readiness level required for the LSST Y1 data:

- *Sample selection*: We have used the specifications from DESC SRD as a guide for the sample characteristics (number density, redshift distribution, associated nuisance parameters) as well as many of the analysis choices (scale cuts, models of nuisance parameters, priors on cosmological parameters). Some of these specifications should be updated given our current knowledge from Stage-III experiments.

- *n(z):* We have used the true redshift information for the source and lens samples in the modeling of this work to decouple the effect of photometric redshift estimation. Future analyses will employ realistic photometric redshift algorithms to determine the impact of potential biases in the estimated redshift distributions.

- *Mask:* We use a very simplistic LSST Y1 mask, without holes and without accounting for the Milky Way region or Rubin's survey strategy. Future analyses should use more realistic masks such as the one from Lochner et al. (2018) or an updated version of that when available. Moreover, we do not account for generally spatially varying systematics, which would usually be corrected with LSS weights. Future analyses should implement and test methods to correct for these effects.

- *Covariance matrix:* Here we rely on a Gaussian covariance and only include mask effects in harmonic space for the CosmoDC2 simulation. Future analyses will include non-Gaussian terms and mask effects both in harmonic and real space.

- *Estimators:* In this analysis we have implemented the standard estimators that have already been used in Stage-III 3×2pt analyses. As an exception, we did not include lens-source clustering effects in the galaxy-galaxy lensing estimator, usually corrected for with the so-called "boost factors", which future analyses should account for. Moreover, a whole suite of alternative estimators optimized for different goals exist in the literature, such as the KNN's (Banerjee & Abel 2021) sensitive to higher order information in the galaxy density field; the COSEBIs (Schneider et al. 2010), an alternative E/B mode decomposition of the cosmic shear information; or the $\Delta\Sigma$ estimator (Sheldon et al. 2004) for galaxy-galaxy lensing, which is often used to perform galaxy-halo connection studies. TXPipe's flexible framework will make it possible to implement and test these estimators in future analyses.

- *Model*: To perform a 3×2pt analysis with the LSST Y1 data set the model will need to be extended in several aspects to *at least* consider: non-Limber terms, Tidal Alignment Tidal Torque (TATT) IA model, redshift-space distortions and magnification effects.

- *Cosmology inference code*: This is the first analysis where FIRECROWN has been used, and as such we have only tested the most basic implementation of it. Future analyses will be able to test more advanced features such as different matter power spectrum models or nuisance parameter marginalization schemes.

7. SUMMARY

In this paper we perform a rigorous validation test on the software pipeline in LSST DESC to be used for a cosmology analysis with three two-point functions: galaxy clustering, galaxy-galaxy lensing, and cosmic shear, or the 3×2pt analysis. The core of the validation surrounds the software package TXPipe, which carries out the measurement of the 3×2pt data vector in both real and harmonic space. But TXPipe also interfaces with several other DESC software packages – CCL (for the theoretical modeling), FIRECROWN (for the cosmological inference), TJPCov (for the covariance matrix calculation), SACC (for storing data vectors and relevant metadata) – which are also tested via this process.

We first validate the pipeline using a Gaussian simulation that is designed to mimic the galaxy sample used in the LSST DESC Science Requirements Document (DESC SRD). In particular, we test the pipeline with the statistical power expected for the first year (Y1) of LSST data. We find that TXPipe can properly recover the input theory to the LSST Y1 precision at both the data vector level and the cosmological constraints, with both the real-space and harmonic-space estimators. This first validation of the pipeline assumes a number of simplifications that will need to be revisited when considering

the first-year data analysis. We leave for future work a validation of the pipeline including additional effects.

We also find that in this set up, the systematics contribute to a significant part of the uncertainties, particularly in the S_8 parameter, which we find it would be $\sim 5(1.8)$ times more tightly constrained under the $\Lambda(w_0-w_a)$CDM model if we fixed shear and redshift calibration parameters. On the other hand we find these systematics only contribute slightly to the uncertainties in w_0 or w_a, the parameters describing the equation of state of dark energy.

We then apply this pipeline to CosmoDC2, a DESC mock galaxy catalog built on an N-body simulation, designed to be sufficiently deep for LSST and have more realistic galaxy characteristics. We find good agreement between the measurements and theory when we only consider scales where we supply the correct theory. We are also able to recover the input cosmological parameters, in particular in the σ_8 and S_8 parameters, which we find are the only parameters that are not prior-dominated in this setup.

Overall, we have developed and validated to LSST-Y1 precision a catalog-to-cosmology framework with DESC tools to obtain cosmological information from weak lensing and galaxy clustering measurements. In this paper we have focused on two-point weak lensing and galaxy clustering estimators, but the scheme we have developed can easily be extended to include higher order estimators and also other probes. Therefore, we have accomplished the first milestone in the roadmap to perform cosmological analyses with LSST data.

7.2. *Affiliations*

[1] University of Chicago, 5801 S Ellis Ave, Chicago, IL 60637, USA

[2] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

[3] Institute of Astronomy, Royal Observatory Edinburgh, University of Edinburgh, Edinburgh EH9 3HJ, United Kingdom

[4] Eidgenoessische Technische Hochschule Zuerich, Rämistrasse 101, 8092 Zürich, Switzerland, Switzerland

[5] Harvard University, Cambridge, MA 02138, USA

[6] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom

[7] Duke University, Durham, NC 27708, USA

[8] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21211, USA

[9] University of California Berkeley, 366 Physics North, MC7300, Berkeley, CA 94720, USA

[10] Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA

[11] Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL 60439, USA

[12] University of Texas, Dallas, 800 W Campbell Rd, Richardson, TX 75080, USA

[13] University of Pennsylvania, Philadelphia, PA 19104, USA

[14] University of Utah, Department of Physics and Astronomy 115 South 1400 East, Salt Lake City, UT 84112-0830, USA

[14] University of Utah, Department of Physics and Astronomy 115 South 1400 East, Salt Lake City, UT 84112-0830, USA

[15] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510-5011, USA

[16] Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-40, Brazil

REFERENCES

Abbott T. M. C., et al., 2018, Phys. Rev. D, 98, 043526
Aihara H., et al. 2018, PASJ, 70, S4
Akeson R., et al. 2019, arXiv e-prints, p. arXiv:1902.05569
Albrecht A., et al. 2006a, arXiv e-prints, pp astro–ph/0609591
Albrecht A., et al. 2006b, arXiv e-prints, pp astro–ph/0609591
Alonso D., et al. 2019, MNRAS, 484, 4127
Amon A., et al. 2022, Phys. Rev. D, 105, 023514
Banerjee A., Abel T., 2021, MNRAS, 500, 5479
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Benítez N., 2000, ApJ, 536, 571
Bridle S., King L., 2007, New Journal of Physics, 9, 444
Brown M. L., et al. 2002, MNRAS, 333, 501
Carrasco Kind M., Brunner R. J., 2014, MNRAS, 438, 3409
Chang C., et al. 2019, MNRAS, 482, 3696
Chisari N. E., et al. 2019, ApJS, 242, 2
Coe D., et al. 2006, AJ, 132, 926
Crenshaw J. F., Doster V., 2022, jfcrenshaw/pzflow: v2.0.7, doi:10.5281/zenodo.6369625, https://doi.org/10.5281/zenodo.6369625
Crocce M., Cabré A., Gaztañaga E., 2011, MNRAS, 414, 329
DES Collaboration 2022a, arXiv e-prints, p. arXiv:2207.05766
DES Collaboration 2022b, Phys. Rev. D, 105, 023520
Dalmasso N., et al. 2020, Astronomy and Computing, 30, 100362
Di Valentino E., et al. 2021, Astroparticle Physics, 131, 102604
Eisenstein D. J., et al. 2005, ApJ, 633, 560
Elvin-Poole J., et al. 2018, Phys. Rev. D, 98, 042006
Elvin-Poole J., et al. 2022, arXiv e-prints, p. arXiv:2209.09782
Fang X., Eifler T., Krause E., 2020, MNRAS, 497, 2699
Flaugher B., 2005, International Journal of Modern Physics A, 20, 3121
Foreman-Mackey D., et al. 2013, PASP, 125, 306
Friedrich O., et al. 2021, MNRAS, 508, 3125
Gatti M., et al. 2020, MNRAS, 498, 4060
Gatti M., et al. 2021, MNRAS, 504, 4312
Gatti M., et al. 2022, Phys. Rev. D, 106, 083509
Giannantonio T., et al. 2008, Phys. Rev. D, 77, 123520
Giblin B., et al. 2021, A&A, 645, A105
Hang Q., et al. 2021, MNRAS, 501, 1481
Hartley W. G., et al. 2020, MNRAS, 496, 4769
Heitmann K., et al. 2019, ApJS, 245, 16
Heymans C., et al. 2021, A&A, 646, A140
Hinton S. R., 2016, The Journal of Open Source Software, 1, 00045
Ivezic Z., Jones L., Lupton R., 2010, Technical Report LSE-40, The LSST Photon Rates and SNR Calculations, v1.2, http://faculty.washington.edu/ivezic/Teaching/Astr511/LSST_SNRdoc.pdf. LSST, http://faculty.washington.edu/ivezic/Teaching/Astr511/LSST_SNRdoc.pdf
Ivezić Ž., et al. 2019, ApJ, 873, 111
Izbicki R., Lee A. B., 2017, Electronic Journal of Statistics, 11, 2800
Jarvis M., Bernstein G., Jain B., 2004, MNRAS, 352, 338
Jeffrey N., et al. 2021, MNRAS, 505, 4626
Kilbinger M., et al. 2017, MNRAS, 472, 2126
Komatsu E., et al. 2011, ApJS, 192, 18
Korytov D., et al. 2019, ApJS, 245, 26
Kovacs E., et al. 2022, The Open Journal of Astrophysics, 5, 1
Krause E., Eifler T., 2017, MNRAS, 470, 2100
Krause E., et al. 2021, arXiv e-prints, p. arXiv:2105.13548
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Laureijs R., et al. 2011, arXiv e-prints, p. arXiv:1110.3193
Leistedt B., Hogg D. W., 2017, ApJ, 838, 5
Levi M., et al. 2019, in Bulletin of the American Astronomical Society. p. 57 (arXiv:1907.10688)
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Li X., et al. 2022, PASJ, 74, 421
Limber D. N., 1953, ApJ, 117, 134
Liu J., et al. 2015, Phys. Rev. D, 91, 063507
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Lochner M., et al. 2018, arXiv e-prints, p. arXiv:1812.00515
Lochner M., et al. 2022, ApJS, 259, 58
Longley E. P., et al. 2023, MNRAS, 520, 5016
MacCrann N., et al. 2022, MNRAS, 509, 3371
Merkel D., 2014, Linux journal, 2014, 2
Miller L., et al. 2007, MNRAS, 382, 315
Muir J., et al. 2020, MNRAS, 494, 4454
Nicola A., et al. 2020, J. Cosmology Astropart. Phys., 2020, 044
Nicola A., et al. 2021, J. Cosmology Astropart. Phys., 2021, 067
Pandey S., et al. 2022, Phys. Rev. D, 106, 043520
Prat J., et al. 2018, Phys. Rev. D, 98, 042005
Prat J., et al. 2022, Phys. Rev. D, 105, 083528
Rau M. M., et al. 2021, Monthly Notices of the Royal Astronomical Society, 509, 4886
Rodríguez-Monroy M., et al. 2022a, MNRAS, 511, 2665
Rodríguez-Monroy M., et al. 2022b, MNRAS, 511, 2665
Schneider P., et al. 2002, A&A, 396, 1
Schneider P., Eifler T., Krause E., 2010, A&A, 520, A116
Secco L. F., et al. 2022, Phys. Rev. D, 105, 023515
Seitz C., Schneider P., 1997, A&A, 318, 687
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Sheldon E. S., et al. 2004, AJ, 127, 2544
Sheldon E. S., et al. 2020, ApJ, 902, 138
Singh S., 2021, MNRAS, 508, 1632
Singh S., et al. 2017, MNRAS, 471, 3827
Springel V., Frenk C. S., White S. D. M., 2006, Nature, 440, 1137
Takahashi R., et al. 2012, The Astrophysical Journal, 761, 152
The LSST Dark Energy Science Collaboration 2018, arXiv e-prints, p. arXiv:1809.01669
Troxel M. A., et al. 2018, MNRAS, 479, 4998
Wright A. H., et al. 2020, A&A, 637, A100

Zhang Z., Becker M. R., Sheldon E. S., 2022a, arXiv e-prints, p. arXiv:2206.07683
Zhang Z. J., et al. 2022b, MNRAS, 514, 2181
Zuntz J., et al. 2015, Astronomy and Computing, 12, 45

Zuntz J., et al. 2018, MNRAS, 481, 1149
Zuntz J., et al. 2021, The Open Journal of Astrophysics, 4, 13
Zürcher D., et al. 2022, MNRAS, 511, 2075
de Jong J. T. A., et al. 2013, The Messenger, 154, 44

APPENDIX

TWO-POINT ESTIMATORS

Harmonic space estimator

We use NaMaster (Alonso et al. 2019) to measure the harmonic space power spectra, based on the pseudo-C_ℓ method. For the harmonic space measurements we start from a map, either of galaxy density δ_g or shear γ. An observed map $\tilde{a}$ is the product of the true map a and a weight map w: $\tilde{a} = wa$. The weights map can either be a binary mask or an inverse-variance local weight, downweighting regions of high noise.

The spherical harmonic coefficients of the observed map $\tilde{a}$ are a convolution of the spherical harmonic coefficients of a and w, given the convolution theorem. This leads to mode-coupling between the power spectra of the observed and true fields and prevents us from directly measuring the desired C_ℓ. We can instead define first the coupled pseudo $\tilde{C}_\ell^{ab}$ between fields a, b with weights v, w respectively:

$$\tilde{C}_\ell^{ab} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \tilde{a}_{\ell m} \tilde{b}_{\ell m}^*, \qquad (A1)$$

which is related to the true power spectrum via the mode-coupling matrix (MCM):

$$\langle \tilde{C}_\ell^{ab} \rangle = \sum_{\ell'} M_{\ell\ell'}^{vw} C_\ell^{ab}, \qquad (A2)$$

where the averaging $\langle \rangle$ is over different realizations of a and b of the initial Gaussian fields. The coupling matrix depends only on the weights of the two fields being correlated. Explicit expressions for the coupling matrix of different combinations of spin-0 and 2 fields can be found in Alonso et al. (2019). If the data covers all sky, then the coupling matrix is the identity. When we are only able to observe a part of the sky, this produces off-diagonal coupling between neighboring multipoles. To obtain C_ℓ^{ab}, we need to invert the coupling matrix. However, $M_{\ell\ell'}^{vw}$ is non-invertible in many scenarios. This is usually solved by binning the power spectra in *bandpowers B*, containing weighted sums of different multipoles:

$$\tilde{C}_q^{ab} = \sum_{l \in q} B_q^\ell \tilde{C}_\ell^{ab}, \qquad (A3)$$

in which case the binned coupling matrix is defined as:

$$\mathcal{M}_{qq'}^{vw} = \sum_{l \in q} \sum_{l' \in q'} M_{\ell\ell'}^{vw}, \qquad (A4)$$

and is usually invertible for sufficiently broad bandpowers. Then, the decoupled $\hat{C}_q^{ab}$ for a chosen bandpower binning q is:

$$\hat{C}_q^{ab} = \sum_{q'} (\mathcal{M}^{vw})_{qq'}^{-1} (\tilde{C}_{q'}^{ab} - \tilde{N}_{q'}^{ab}), \qquad (A5)$$

where $\tilde{N}_{q'}^{ab}$ is the (binned) mode-coupled noise power spectrum, which needs to be removed first for autocorrelations. The form of the noise power spectrum depends on the maps being correlated, and their calculation is described below in A.1.1. The results of Equation (A5) are the measurements we plot in Figures 4, 5 and 6.

Then, the decoupled $\hat{C}_q^{ab}$ would be an unbiased estimator of the true power spectrum evaluated at, e.g. the central multipole of each bandpower, if the power spectra was exactly constant within each bandpower. Since that is not the case, when comparing with theory we need to propagate this averaging with the *bandpower window functions*, which relate the theoretical prediction for the bandpowers C_q^{ab} to the theory power spectrum C_ℓ^{ab}:

$$C_q^{ab} = \sum_\ell \mathcal{F}_{q\ell}^{vw} C_\ell^{ab}, \qquad (A6)$$

where

$$\mathcal{F}_{q\ell}^{vw} = \sum_{q'} (\mathcal{M}^{vw})_{qq'}^{-1} \sum_{\ell' \in q'} w_{q'}^{\ell'} M_{\ell\ell'}^{uv}. \qquad (A7)$$

In Figure 11 we show the impact of this effect with the angular binning that we use in this work for harmonic space, which are 17 logarithmically spaced bandpowers between $\ell = 20$ and $\ell = 3722$, to match the scales defined in the SRD. We find small ℓ values are impacted the most, reaching a $\sim 20\%$ effect at the largest scale. We use a common HEALPIX resolution of $N_{\rm side} = 4096$ for all sky maps.

Finally, we apply a beam correction to account for the window function of `HEALPix`, which depends on $N_{\rm side}$. The correction is largest for small scales (large ℓ), reaching $\sim 3\%$ at $\ell \simeq 3000$.

Harmonic space noise subtraction

In Equation A5 we subtract the mode-coupled noise power spectrum $\tilde{N}_{q'}^{ab}$ from the measured cosmological signal $\tilde{C}_{q'}^{ab}$. In TXPIPE we have implemented two different ways to compute the noise power spectrum, analytically and using a number of realizations of noise maps. Here we detail these implementations for the galaxy clustering and cosmic shear estimators in harmonic space.

For galaxy clustering, the noise is due to poisson sampling of the fields when we observe galaxies. We can compute the *coupled* noise analytically by averaging the mask and dividing by the number density of galaxies n_g, in each redshift bin:

$$\tilde{N}_{\rm coupled,\ clustering} = \frac{\langle {\rm mask} \rangle}{n_g}, \qquad (A8)$$

where in our case the mask is uniform with unity values within the footprint and with zero values outside of it, i.e. $\langle {\rm mask} \rangle$ will approximately be the same as $f_{\rm sky}$[29].

---

[29] Note that the decoupled noise can be expressed as: $N_{\rm decoupled,\ clustering} \simeq \frac{1}{n_g}$, and that the decoupling operation corresponds to approximately an $f_{\rm sky}$ factor.
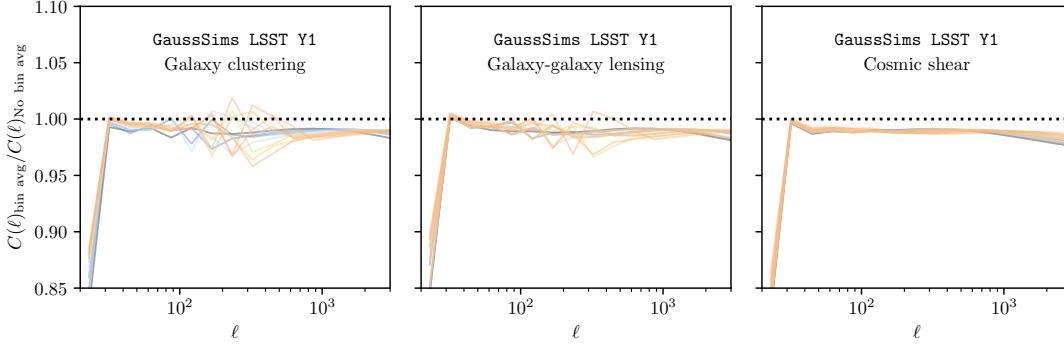
FIG. 11.— Comparison between the theory data vector in harmonic space including proper bin averaging as described in App. A using Equation (A6) vs. using the equations described in Section 2 evaluated at the center ℓ of each angular bin. The colors represent different redshift bins, with the darker (black) colors corresponding to lower redshift, and the lighter (orange) colors to higher redshift bin combinations.

For the cosmic shear case, the coupled noise can be computed as the *sum of weights* estimator from Equation (2.24) in Nicola et al. (2021):

$$\tilde{N}_{\text{coupled, cosmic shear}} = \left\langle \sigma_{e,\text{pix}}^2 \right\rangle A_{\text{pix}}, \text{ for } \ell \geq 2, \qquad \text{(A9)}$$

and zero for $\ell < 2$ (in general $\tilde{N} = 0$ for $\ell < \text{spin}$), where A_{pix} is the area of each pixel, the averaging $\langle \rangle$ happens across all pixels in the sphere and $\sigma_{e,\text{pix}}^2$ is the sum of the weighted variance of the calibrated ellipticity e computed in each pixel as

$$\sigma_{e,\text{pix}}^2 = \frac{1}{N} \sum_{i}^{N} \frac{1}{2} \left(e_{1,i}^2 + e_{2,i}^2 \right) w_i^2, \qquad \text{(A10)}$$

where w_i are the weights associated with the source sample for each object i in a given pixel that has a total of N galaxies. $\sigma_{e,\text{pix}}^2 = 0$ outside the footprint. For Cos-moDC2, we calibrate this quantity under the Meta-calibration framework with $\sigma_{e,\text{pix}}^2 = R_{\text{inv}}^2 \sigma_{e(\text{uncalibrated}),\text{pix}}^2$, where in this case we only use the diagonal part of the mean of the response matrix.

We cross-check the above estimates with an alternative method based on generating a given number of noise realizations. For this work we use 30 realizations of the noise maps. For cosmic shear, we generate maps with random rotations of the ellipticities, and then compute the pseudo-C_ℓ in each noise realization r, and estimate the noise as the mean across all of them, following the method from Nicola et al. (2021). For galaxy clustering, we generate gaussian realization of the density field, and then for each noise map we measure the pseudo-C_ℓ for half the galaxies in the map, which we label as $h1$ and $h2$. Then, we subtract the measurements from each half to null the signal and be left with the noise part, which we divide by 4 to account for the fact that the number density for each split is $1/2$ of the total one:

$$\tilde{N}_{\text{coupled, clustering, maps}} = \frac{1}{4} \left\langle \tilde{C}_\ell^{h1} - \tilde{C}_\ell^{h2} \right\rangle_r. \qquad \text{(A11)}$$

We find agreement between the analytic and the noise maps realizations methods, which we show in Figure 12 for the CosmoDC2 simulation.

Real space estimator

For the real space measurements we start from galaxy or shear catalogs and not from maps as in the harmonic space case. We use $\xi_\pm$ as the estimators for cosmic shear, defined in terms of the tangential (t) and cross ($\times$) components of the ellipticity $\hat{e}$ defined along the line that connects each pair of galaxies a,b. Correlating galaxies in a pair of redshift bins (i,j) we define,

$$\xi_\pm^{ij}(\theta) = \frac{\sum_{ab} w_a w_b \left(\hat{e}_{t,a}^i \hat{e}_{t,b}^j \pm \hat{e}_{\times,a}^i \hat{e}_{\times,b}^j \right)}{\sum_{ab} w_a w_b}, \qquad \text{(A12)}$$

with inverse variance weighting w and where the sums run over pairs of galaxies a,b, for which the angular separation falls within the range $|\boldsymbol{\theta} - \Delta\boldsymbol{\theta}|$ and $|\boldsymbol{\theta} + \Delta\boldsymbol{\theta}|$. The ellipticities that enter Equation (A12) are corrected for residual mean shear, such that $\hat{e}_k^i \equiv e_k^i - \langle e_k \rangle_i$ for components $k \in (1,2)$ and redshift bin i. Also, they have already been corrected by the Metacalibration response factors as described below.

For galaxy-galaxy lensing, the mean tangential shear estimator is usually expressed as:

$$\hat{\gamma}_t^{ij}(\theta) = \frac{\sum_{LS} w_{LS} \hat{e}_{t,LS}^{ij}(\theta)}{\sum_{LS} w_{LS}(\theta)} - \frac{\sum_{RS} w_{RS} \hat{e}_{t,RS}^{ij}(\theta)}{\sum_{RS} w_{RS}(\theta)}, \qquad \text{(A13)}$$

where LS refers to lens-source galaxy pairs that are separated by a given angular scale that falls within the bin. RS refers to random-source pairs, and this term removes the tangential shear around a sample of random points to correct for mask effects and reduce the noise in the jackknife covariance[30] (Singh et al. 2017; Prat et al. 2018). Note in this work we do not include the effect of lens-source clustering in the tangential shear estimator, usually referred as "boost factors", which is left for future work. In our setup we do not expect them to have a significant impact given the scale cuts we use and the mostly non-overlapping redshift bins between lenses and sources.

Finally, for the angular 2-point correlation function, $w(\theta)$ the typical estimator is the Landy-Szalay estimator

---

[30] We do not use the jackknife covariance in this work to infer cosmology but nonetheless we validated its implementation in TXPipe, which is useful for other purposes.
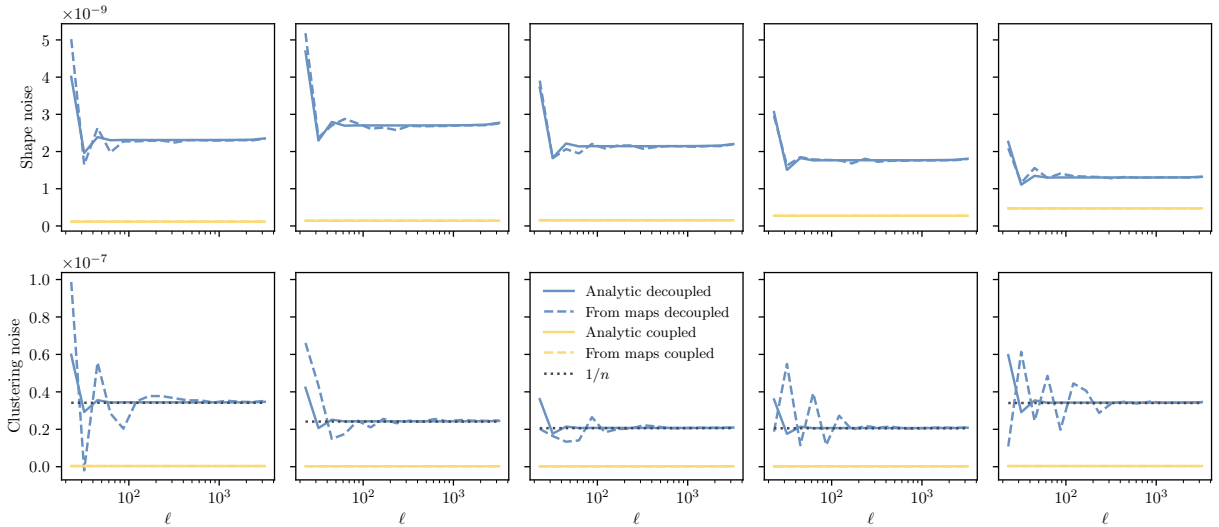
FIG. 12.— Noise power spectrum for COSMODC2, computed by TXPIPE as described in Section A.1.1.

(Landy & Szalay 1993) which can be written as:

$$\hat{w}^{ij}(\theta) = \frac{D^i D^j - D^i R^j - D^j R^i + R^i R^j}{R^i R^j} , \qquad (A14)$$

where DD, RR and DR are the normalized galaxy-galaxy, random-random and galaxy-random pair counts within an angular bin, respectively. We use 20 log-spaced θ bins between 2.5 and 250 arcminutes for all the probes in real space.

Pixel based estimators

Both the galaxy-galaxy lensing and the galaxy clustering estimators written above require a sample of random points. This sample usually needs to be at least 20 times larger than the sample of lens galaxies (Prat et al. 2022), to avoid adding significant extra noise to the measurements[31]. We found that for the Gaussian simulation built following the LSST Y1 SRD lens number densities and area, it becomes infeasible to apply these estimators due to the high amount of memory they require, even after attempting several memory optimizations. As a result, we also implement a pixel-based estimator which we use for the fiducial measurements presented in Section 4.3. The pixel-based version of the Landy-Szalay estimator between two lens redshift bins i, j is

$$\hat{w}^{ij}(\theta) = \sum_{l=1}^{N_{\mathrm{pix}}} \sum_{k=1}^{N_{\mathrm{pix}}} \frac{(N_l^i - \bar{N}^i)(N_k^j - \bar{N}^j)}{\bar{N}^i \bar{N}^j} \Theta_{l,k}, \qquad (A15)$$

where N_l is the galaxy number density in pixel l, $\bar{N}$ is the mean galaxy number density over all pixels within the footprint and $\Theta_{l,k}$ is a top-hat function which is equal to 1 when pixels l and k are separated by an angle θ within the angular bin. In the limits of an infinitely large random catalog and small enough pixel size this is equivalent to the Landy-Szalay estimator. The fractional

[31] We tested using less amount of random points, in particular 12 times larger than the lens sample, and found this was insufficient to obtain a good fit to the theoretical predictions.

coverage of each pixel is taken into account in the calculation of N_l and $\bar{N}$. We also use TREECORR for this pixel-based estimator. As input for this function, we use the same density contrast maps that are used in the harmonic space two-point measurements. Note this estimator has already been implemented with DES Y3 data in Rodríguez-Monroy et al. (2022b). The analogous pixel-based estimator for the mean tangential shear can be written as:

$$\hat{\gamma}_t^{ij}(\theta) = \sum_{l=1}^{N_{\mathrm{pix}}} \sum_{k=1}^{N_{\mathrm{pix}}} \frac{(N_l^i - \bar{N}^i)}{\bar{N}^i} \gamma_t^j \Theta_{l,k} \qquad (A16)$$

where we can effectively use the correlation using the density contrast and ellipticity maps, named KG as implemented in the TreeCorr code. We verify that the difference between this pixel version of the estimator and that using random points is negligible for the angular scales we consider given the nside = 4096 we use and the scales cuts that are applied for the galaxy clustering and galaxy-galaxy lensing probes.

Response factors

We use response factors that account for shear and selection biases within the METADETECTION framework introduced in Sheldon et al. (2020). In all the above equations, the shear per each galaxy $\hat{e}$ has already been corrected by the response factors. In particular, we use the full response matrix to correct for the shear. We measure the response matrix in each of the source redshift bins, for the whole ensemble of galaxies in each bin, and then apply this mean response factor to each of the galaxies individually, before using the two-point estimators as described above. We test this implementation in TXPipe with the COSMODC2 measurements, for which we generate a non-unity response factor. While this implementation is already more advanced than what has been used in Stage–III surveys (e.g as in Amon et al. 2022; Secco et al. 2022; Prat et al. 2022), more sophisticated treatments may be required for LSST estimates, such as computing response factors for smaller ensembles or the deep-field approach described in Zhang et al.

(2022a).

MASK EFFECT ON THE COVARIANCE

In Figure 13 we show the impact of including the mask effects on the covariance for harmonic space. We compare the covariance obtained with the f_{sky} approximation with the one where the coupling between modes due to the mask is taken into account. We compute the coupling between modes with the DESC package TJPCov, which calls the NaMaster code. We find the impact is biggest for the largest scales, corresponding to lower ℓ. In particular, for our simple LSST Y1 mask shown in Figure 2 we find that for the largest scale used in this work of $\ell = 20$ the effect is $\sim 15\%$ on the square root of the diagonal elements of the covariance. We find this to be the case for all the probes. At small scales the impact of the mask is $\leq 5\%$. We check the impact on the cosmological contours is negligible in Figure 14, and therefore do not include it in our fiducial covariance with the LSST Y1 Gaussian simulation. We note that these conclusions might differ when considering a more realistic mask (e.g. masking bright stars will affect all angular scales, and not only large scales).

On the other hand, we find this effect to be much larger for the smaller ~ 440 deg.^{2} CosmoDC2 mask, where we find it can be as large as $\sim 200\%$ for $\ell = 20$. We note however that $\ell = 20$ is quite close to the edge of the footprint, which we would normally not use in a typical cosmological analysis. We also find that the difference in the two covariances is most significant for the cosmic shear probe. Given the magnitude of this effect, we use the covariance including mask effects for our fiducial harmonic space results on CosmoDC2. We do not find the impact of the mask to be as important for the real space counterpart, since there we do not use such large scales: for the real space analysis we use scales up to 250 arcmin, which corresponds to about $\ell \sim 45$, using the $\ell \sim \pi/\theta$ approximate scaling.

COMPARISON WITH DESC SRD FORECASTS AND STAGE-III SURVEYS

In this appendix we compare our results on the LSST Y1 Gaussian simulation with the forecast from the DESC SRD and from Stage III surveys. However, neither of those are perfect comparisons since there are some differences in the analyses. First we list the differences between our analysis and the DESC SRD forecast:

1. **Input cosmology:** The width of the priors is matched to the DESC SRD forecast one but not their center values, since we wanted to use the same cosmological parameters that were input to CosmoDC2.

2. **Covariance:** The DESC SRD uses a non-Gaussian covariance, while we only use the Gaussian terms.

3. **Noise:** The DESC SRD used noiseless datavectors while we use a noisy realization of the Gaussian simulation.

4. **Intrinsic alignment model:** The DESC SRD used a 4-parameter IA model, with an overall redshift scaling parameter with a Gaussian prior with mean $\mu = 5$, $\sigma = 3.9$ (while we use the same σ, but center

it at zero, since the simulations we use do not include intrinsic alignment effects); a power-law luminosity scaling parameter (which we do not include), a redshift scaling parameter with $\mu = 0$ and $\sigma = 2.3$ (identical as ours) and an additional high-redshift scaling parameter (not included here).

Thus, the comparison with the DESC SRD is not completely straightforward but it is still useful as a validation, since each analysis uses very different software tools. Note that we only compare the results in the w_0-w_aCDM model, since that is the only one that was considered in the DESC SRD. We display the mean and 1-σ uncertainties for both cases in Table 7 and plot the 1D posteriors in Figure 15. Note that for the plot we have shifted the DESC SRD posteriors to match their mean to the input values of the Gaussian and CosmoDC2 simulations that we use in this work, for an easier visual comparison. Generally we find that the constraining power of our posteriors is similar to what was predicted in SRD in most of the parameters. We only obtain significant differences in the parameters describing the equation of state of dark energy: we find the 1-σ uncertainties in w_0 are ~ 1.3 times more constraining than originally predicted and ~ 2 times more constraining for w_a. A potential explanation for these differences is that we assume a simpler IA model (with 2 free parameters instead of the 4 in the SRD).

Then, we also compare the results from this work with Stage-III posteriors, in particular with the DES Y3 3×2pt w_0-w_aCDM analysis from the extensions work in DES Collaboration (2022a), which assumes the same 2-parameter (NLA) IA model as we do, and allows us to perform a reasonable comparison. However, an important difference between the analyses are the priors: in our work we use the same priors that were used in the DESC SRD, which are Gaussian and are informative in some of the cosmological parameters, such as the Hubble constant, Ω_b and w_a, as seen in Fig. 15 comparing the yellow and the gray posteriors, while in the DES Y3 analyses all the priors are flat and non-informative in the cosmological parameters (i.e. it would look flat in all the panels from Fig. 15). With these caveats in mind, we find that the LSST Y1 results on the Gaussian simulation provide a constraint in S_8 that is 2.5 times more constraining than DES Y3 and a significant improvement in the rest of the cosmological parameters. Interestingly, we also find that a LSST Y1 3×2pt analysis will be able to place constraints in the tilt of the power spectrum n_s, which was completely unconstrained in Stage-III surveys analyses.

FULL PARAMETER POSTERIORS

Here we report the rest of the 2D parameter posteriors for the LSST-Y1 like Gaussian simulation. In Figure 16 we show the cosmological parameters, for ΛCDM and w_0-w_aCDM, in comparison with the prior. We find that all the parameters are constrained with respect to the prior, not only the usual Ω_m and σ_8 parameters. In particular, we obtain a relatively tight constraint on n_s, observed for the first time in 3×2pt analyses.

We also show the difference in the posteriors when we fix the nuisance parameters controlling the redshift and shear calibration parameters (but still varying the in-
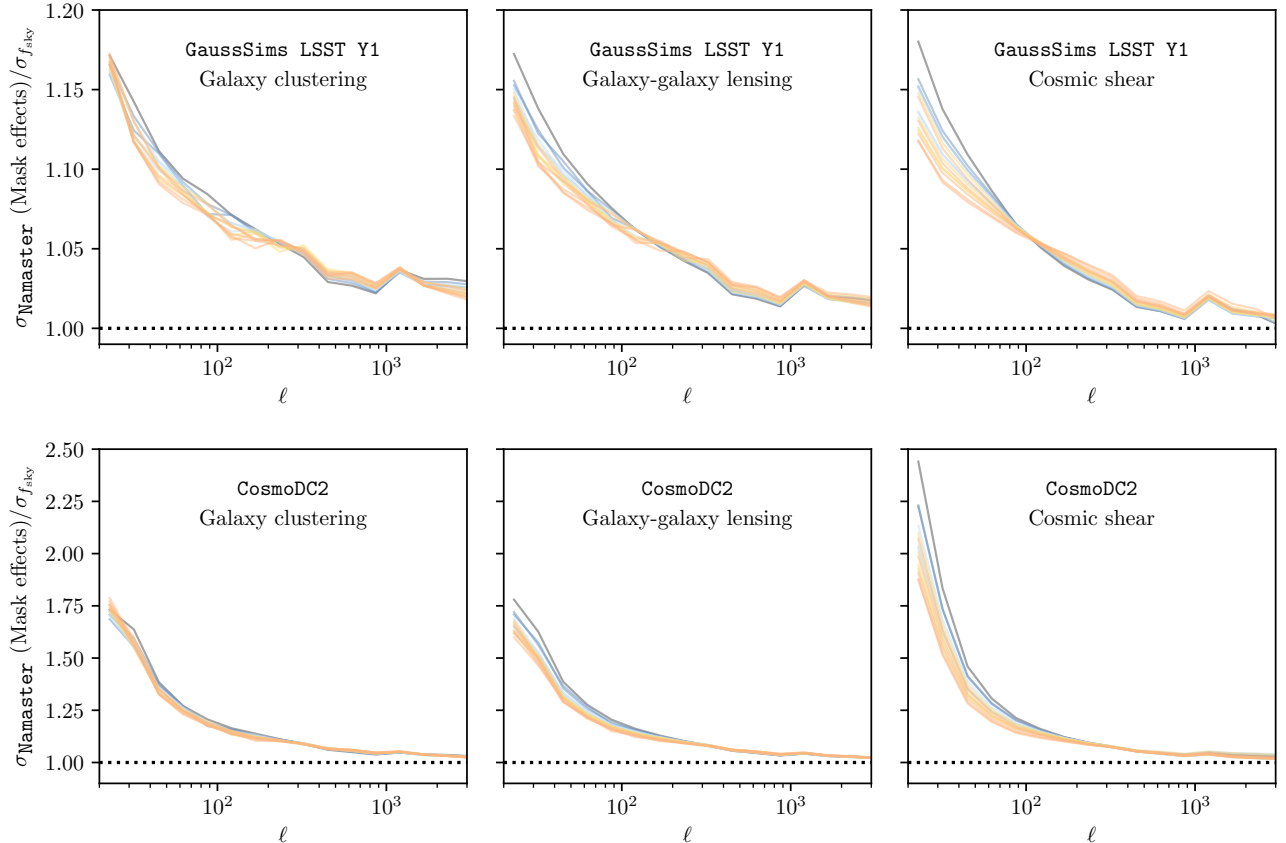
FIG. 13.— Comparison between the covariance including coupling between the ℓ modes due to mask effects and the one using the $f_{\rm sky}$ approximation. The covariance that includes mask effects is computed with the DESC code `TJPCov` that is based on `NaMaster`. The colors represent different redshift bins, with the darker (black) colors corresponding to lower redshift, and the lighter (orange) colors to higher redshift bin combinations.

trinsic alignment and galaxy bias parameters). We find that fixing these systematics results in S_8 constraints that are ~5 times more precise in ΛCDM, both in real space and harmonic space. In w_0-w_aCDM, S_8 uncertainties decrease ~1.8 times when fixing the same observational systematics, ~1.3 times for w_0 and do not change significantly for w_a. We note that the DESC SRD only tried to optimize the w_0 and w_a parameters in the w_0-w_aCDM model and thus it is somewhat expected that we find that S_8 is systematics dominated, also especially in the ΛCDM model which was not considered in the DESC SRD.

We have also checked the posteriors of the nuisance parameters, for the IA, shear and redshift calibration parameters. In Figure 18 we show the IA parameter posteriors, finding that the amplitude of intrinsic alignments has a tight posterior while the parameter controlling its scale dependence is barely constrained. We show the posteriors for the multiplicative shear bias parameters in Figure 17. We find they are being self calibrated by the data themselves with a substantial improvement with respect to the prior, especially at higher redshift. The redshift calibration parameters are shown in Figure 17 and also present significant self-calibration, especially for the lens redshift bins. Overall we find that most of the nuisance parameters are self-calibrated by the data and have tighter posteriors than the prior. This is interesting, and partly due to the fact that the DESC SRD requirements on the redshift and shear calibration pa-

rameters are rather conservative (similar to current values in Stage-III experiments). While we only show the posteriors for ΛCDM we have checked the ones for the w_0-w_aCDM model, which exhibit a very similar behavior.

COSMODC2 DATAVECTORS

In Figures 19, 20 and 21 we display the measurements for cosmic shear, galaxy-galaxy lensing and galaxy clustering, respectively, for the CosMoDC2 simulation. Visually inspecting the measurements, we observe that the large scales suffer from some sharp features below $\ell \sim 50$, which probably indicates that such large scales should be used carefully given the relatively small area of CosMoDC2. We also observe some apparent deviations between the theory and the measurements for the real space galaxy clustering case. However, we note that these measurements are highly correlated and that we find good agreement between the theory and the measurements within the scale cuts using the χ^2 and PTE metrics described in Section 4 and shown in Table 8.

AFTERMATH OF THE CODE COMPARISON WITH DESCQA AND VALIDATION PROCESS

This appendix complements the discussion from Section 6, where we mention that comparing different pieces of the output with external codes often exposes problems
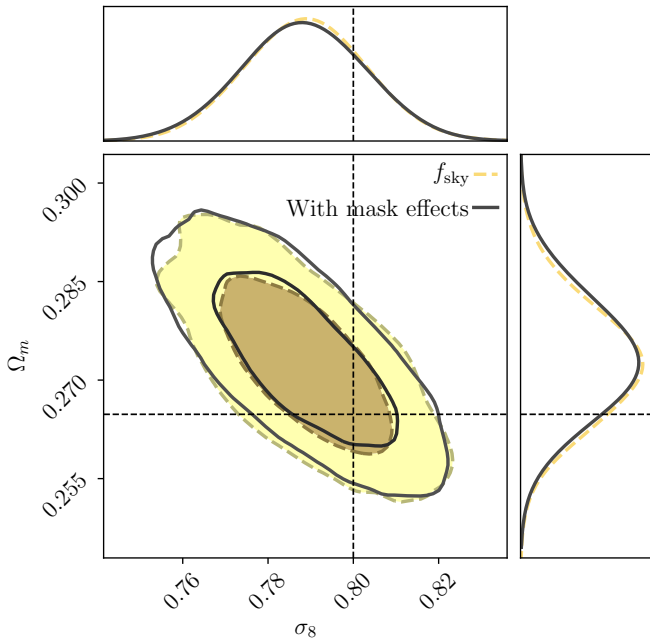
Fig. 14.— Comparison at the cosmological posteriors level for the LSST Y1 like Gaussian simulation in the ΛCDM model between using the covariance that includes coupling between the ℓ modes due to mask effects and the one using the f_{sky} approximation, using the harmonic space datavectors.

that would otherwise be hard to track. Here we illustrate this point by giving examples of the aftermath of such a comparison, in particular of the one comparing the real-space two-point correlation function measurements and theory data vectors from TXPIPE's output to the ones from another DESC code, DESCQA[32], previously validated in Kovacs et al. (2022). From this comparison we identified:

- A bug when constructing the redshift distributions in TXPIPE related to an offset in the redshift binning vs. the histogram, by half a bin. This produced noticeable differences in the theory data vectors yielded by each code.

- A bug in the mean shear subtraction. The mean shear across the whole sample per each redshift bin was not being subtracted correctly.

- A bug related to a mislabeling/typographical error. A name mismatch in the configuration file made us think we were using `bin_slop=0` (a parameter in `TreeCorr` that when non-zero introduces some randomness in the result while speeding up the calculation), when in reality this option was not being transferred to `TreeCorr`.

Other examples of issues that were identified in the process of validating the two-point measurements and theory data vectors for the Gaussian simulations include:

- Bugs related to the ingestion of the input catalogs: Initially the catalog that was fed in to TXPIPE was already masked, and the catalog level mask was

---

[32] https://github.com/LSSTDESC/descqa

interacting with the mask applied within TXPIPE in unexpected ways.

- Bugs related to using patches in several stages, which are needed to handle such large catalogs. For instance, clashing cache name directories produced identical outputs from different inputs, which was identified and then fixed by making a cache directory based on creation time with an absolute path.

- Some redshift distributions produced by TXPIPE were susceptible to numerical errors when being integrated to produce theory predictions, which led to incorrect theory data vectors.

This paper was built using the Open Journal of Astrophysics LaTeX template. The OJA is a journal which provides fast and easy peer review for new papers in the `astro-ph` section of the arXiv, making the reviewing process simpler for authors and referees alike. Learn more at http://astro.theoj.org.
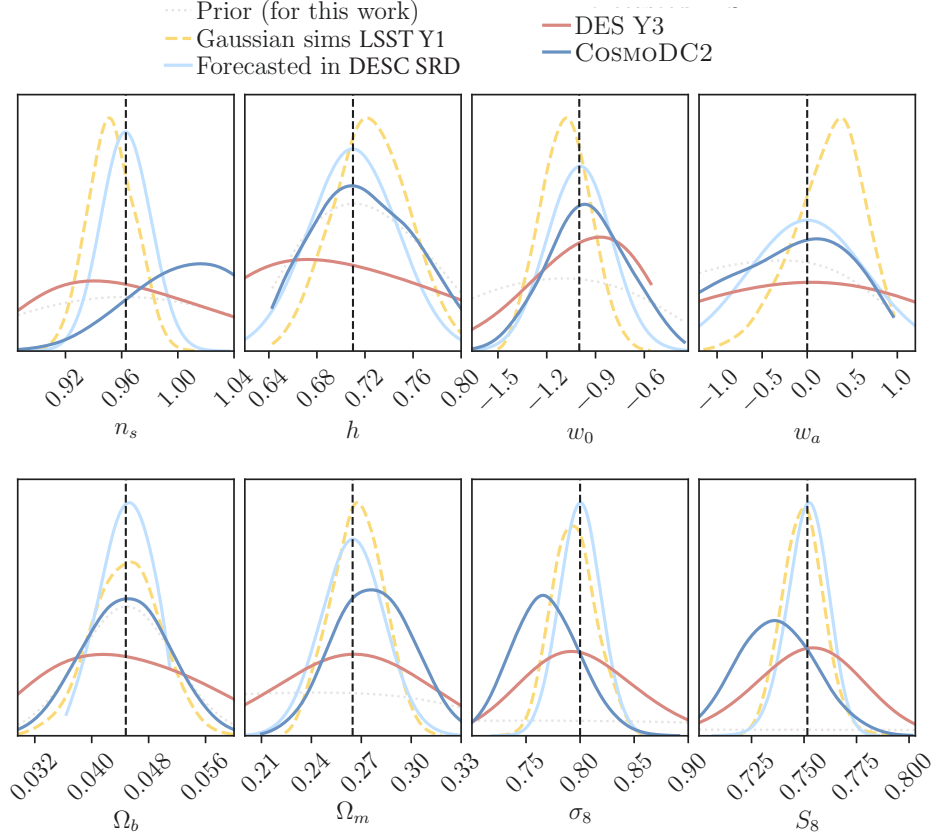
FIG. 15.— Comparison of the 1-D posteriors from different data sets under the $w_0 - w_a$CDM model. First, we compare the results obtained in this paper from the 440 deg.^{2} CosmoDC2 simulation and the 12300 deg.^{2} LSST-Y1 like Gaussian simulation with the prior. Most of the parameters are prior dominated for CosmoDC2. The vertical dashed lines correspond to the input values for both of these simulations. Then, we compare these results with the forecasted predictions from the DESC SRD (shifted in this plot to match the same values as CosmoDC2). We also compare the constraining power to that of stage-III surveys, in particular to the DES Y3 3×2pt $w_0 - w_a$CDM results from the extensions work from DES Collaboration (2022a), with the caveat that they assume different priors with respect to this work (they assume flat priors in all cosmological parameters). The mean contours of DES Y3 are also shifted to match the input of CosmoDC2.
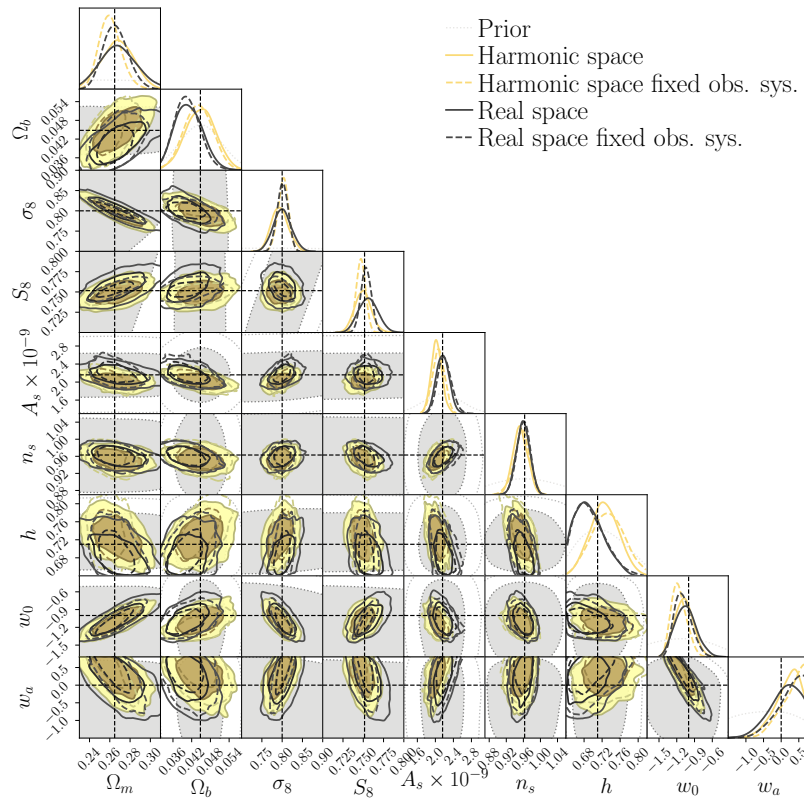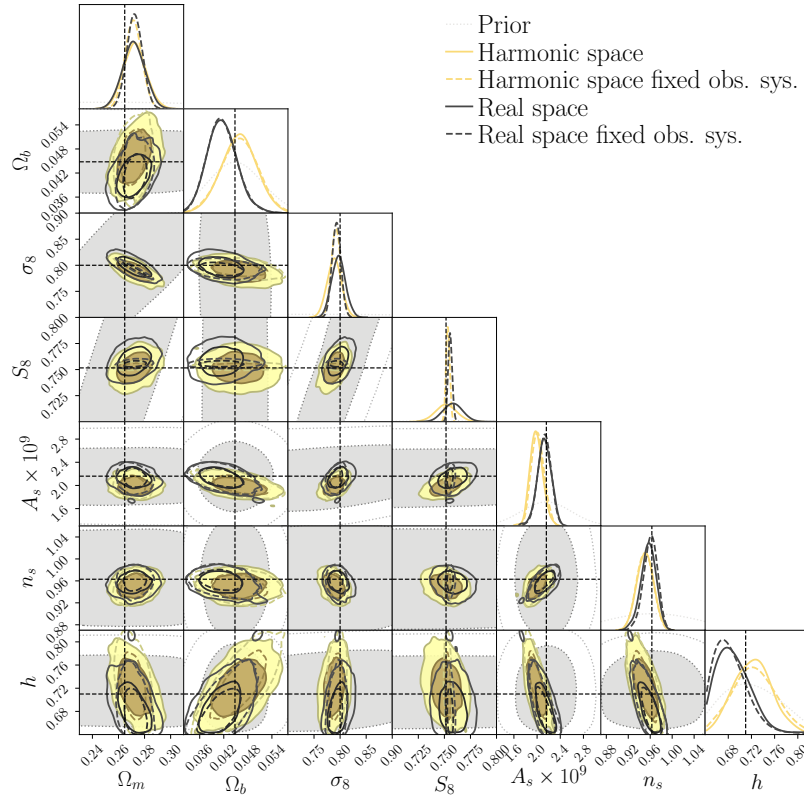
FIG. 16.— ΛCDM (*top*) and w_0-w_aCDM (*bottom*) cosmological contours for the LSST-Y1 like Gaussian simulation.
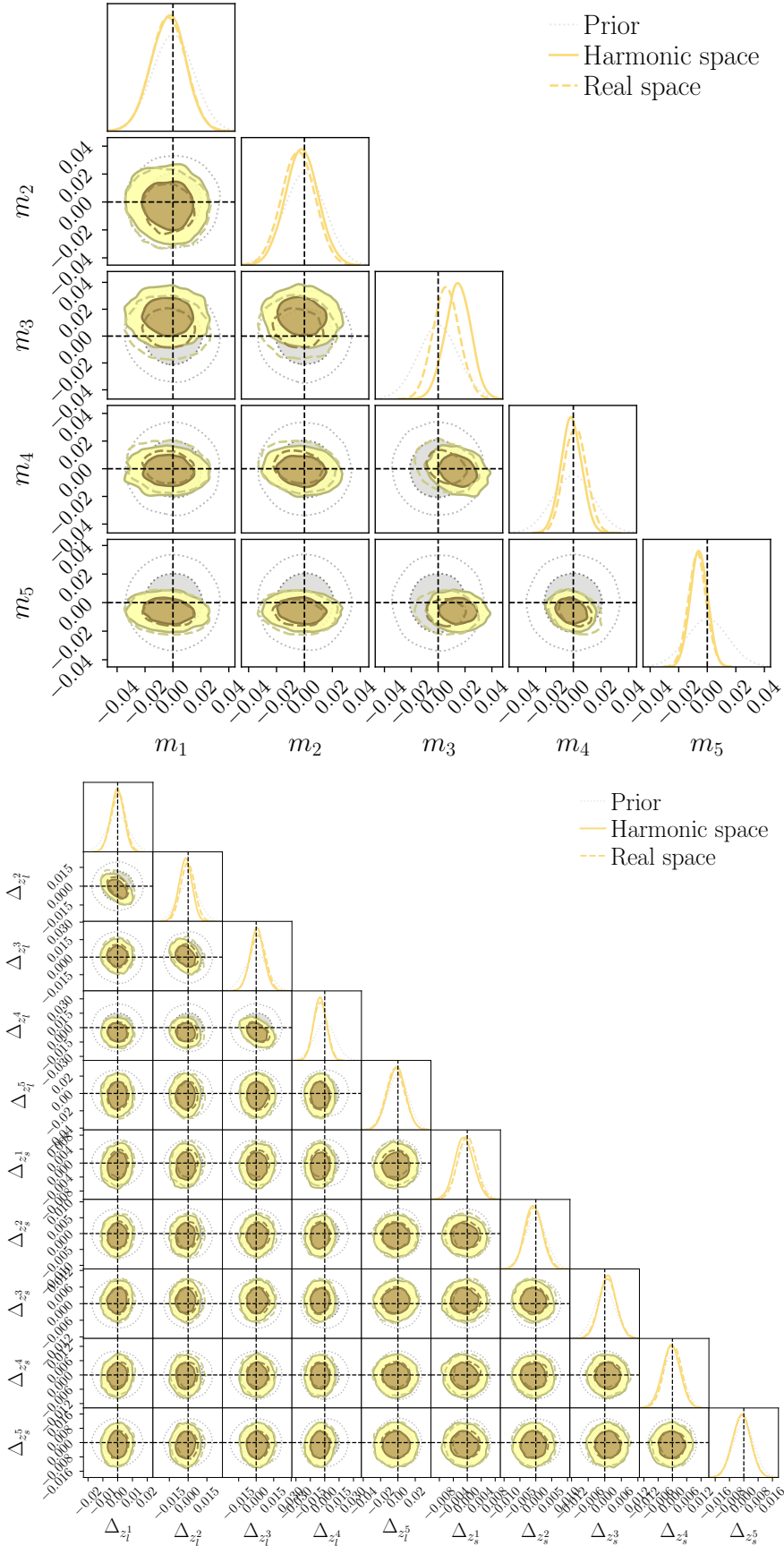
FIG. 17.— Multiplicative shear bias posteriors (top) and redshift calibration parameters (bottom) for the LSST-Y1 like Gaussian simulation assuming the ΛCDM model.
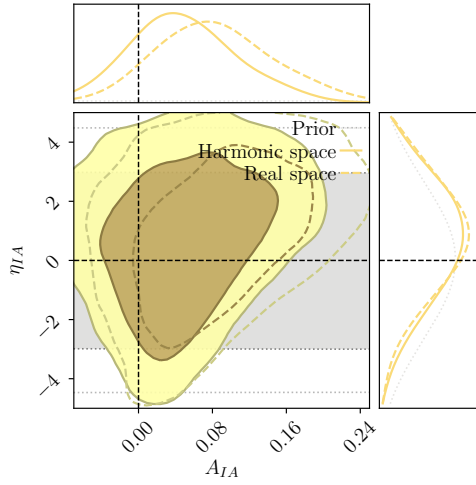
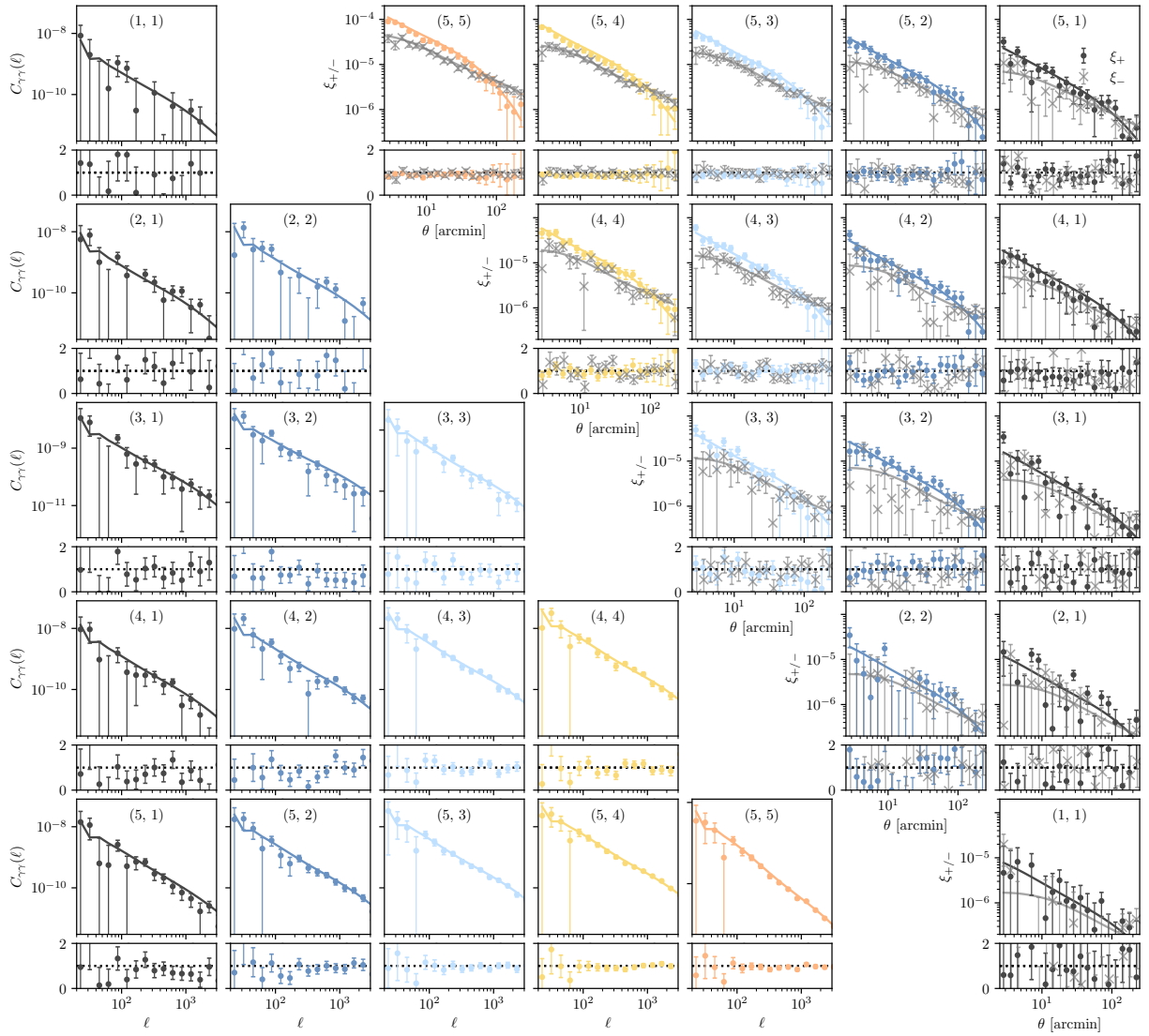FIG. 18.— Intrinsic alignment parameters posteriors for the LSST-Y1 like Gaussian simulation assuming the ΛCDM model.

FIG. 19.— Analogous figure to Fig. 4 for the CosmoDC2 simulation.
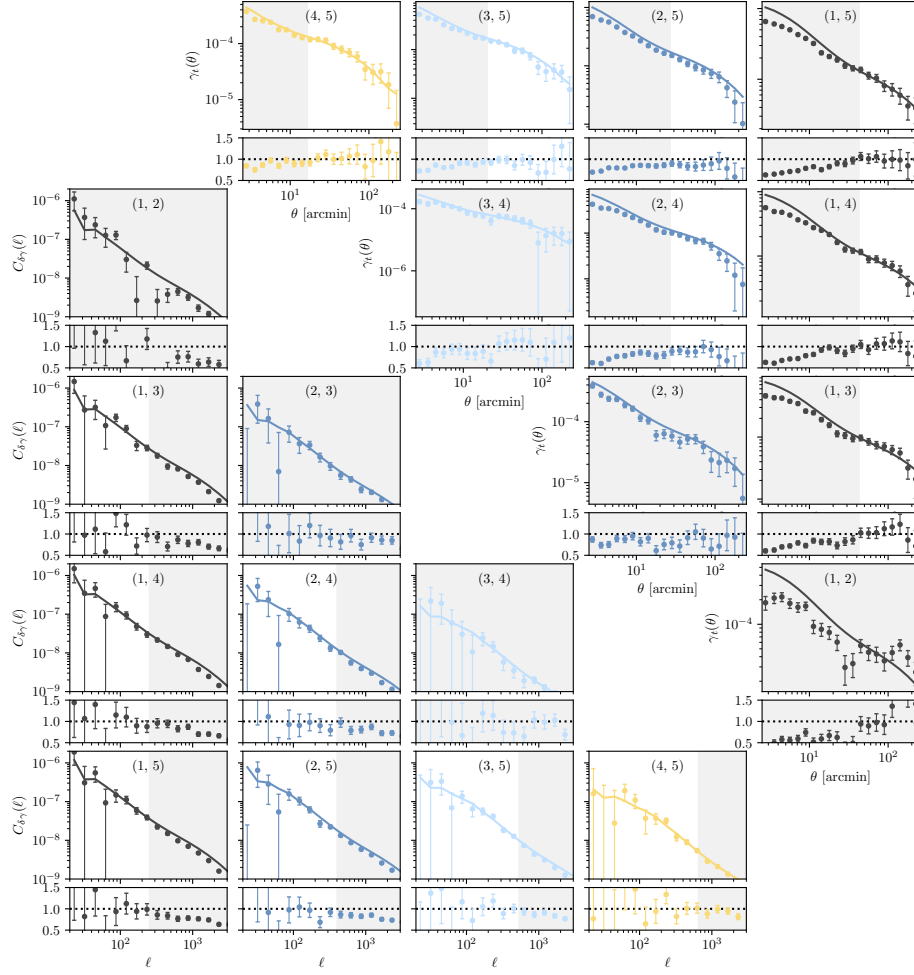
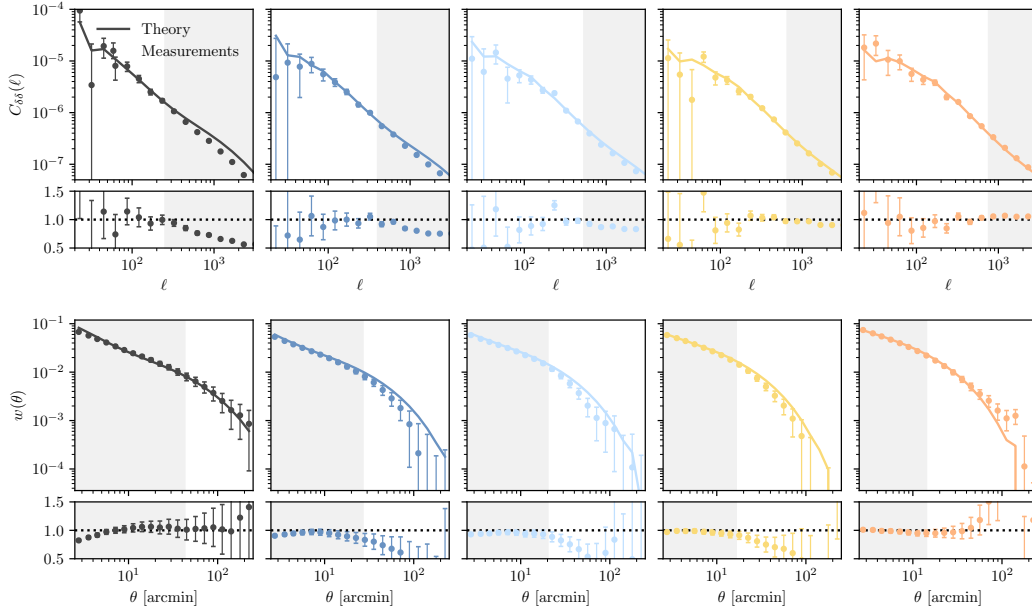FIG. 20.— Analogous figure to Fig. 5 for the COSMODC2 simulation.

FIG. 21.— Analogous figure to Fig. 6 for the COSMODC2 simulation. Note that these measurements are highly correlated and that we find good agreement between the theory and the measurements using the PTE metric as shown in Table 8.