

Constraining power of wavelet vs power spectrum statistics for CMB lensing and weak lensing with learned binning

Kyle Boone^{1,*}, Georgios Valogiannis^{2,3,†}, Marco Gatti^{3,‡} and Cora Dvorkin^{1,§}

¹*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

²*Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA*

³*Kavli Institute for Cosmological Physics, Chicago, Illinois, 60637, USA*



(Received 22 October 2025; accepted 25 February 2026; published 16 March 2026)

We present forecasts for constraints on the matter density (Ω_m) and the amplitude of matter density fluctuations at $8h^{-1}$ Mpc (σ_8) from cosmic microwave background lensing convergence (κ_{CMB}) maps and galaxy weak lensing convergence (κ_{WL}) maps. For κ_{CMB} auto statistics, we compare the angular power spectra (C_ℓ 's) to the wavelet scattering transform (WST) coefficients. For $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ statistics, we compare the cross angular power spectra to wavelet phase harmonics (WPH). This work also serves as the first application of WST and WPH to these probes. For κ_{CMB} , we find that WST and C_ℓ 's yield similar constraints in forecasts for all surveys considered in this work. When κ_{CMB} is crossed with κ_{WL} projected from *Euclid* Data Release 2, we find that WPH outperforms cross- C_ℓ 's by factors between 2.2 and 3.4 for individual parameter constraints. To compare these different summary statistics, we develop a novel learned binning approach. This method compresses summary statistics while maintaining interpretability. We find this leads to improved constraints compared to more naive binning schemes for our wavelet-based statistics, but not for C_ℓ 's. By learning the binning and measuring constraints on distinct datasets, our method is robust to overfitting by construction.

DOI: [10.1103/ccs8-b99y](https://doi.org/10.1103/ccs8-b99y)

I. INTRODUCTION

Gravitational lensing of photons occurs when light is deflected by a mass distribution between the source and the observer [1]. By observing distortions in the shapes of numerous galaxies from weak lensing (hereafter WL), statistical methods can be used to obtain integrated line-of-sight mass maps (e.g., Ref. [2]). Current and upcoming surveys, such as the *Euclid* mission [3], the Dark Energy Survey ([4]), the Kilo-Degree Survey ([5]), the Hyper Suprime-Cam ([6]), and the *Vera C. Rubin* Observatory's Legacy Survey of Space and Time ([7]), will measure galaxy shapes, enabling measurements of these integrated mass maps. Additionally, by observing the lensing of the cosmic microwave background (hereafter CMBL), one can obtain mass maps integrated all the way back to the surface of last scattering (e.g., Ref. [8]). Examples of surveys that perform this measurement include *Planck* [9], the Atacama Cosmology Telescope (ACT) [10], the South Pole Telescope (SPT) [11], and the *Simons Observatory* (SO) [12]. Lensing-convergence maps from these surveys are used to constrain cosmology by comparing measured

summary statistics with predictions from analytic theories or forward-modeled simulations, within either a likelihood-based or simulation-based inference framework.

If the convergence maps were Gaussian random fields, all of their information would be captured by two-point statistics. However, due to nonlinear gravitational evolution, the large-scale structure of the universe has non-Gaussianities that become more pronounced at low redshift. Since two-point statistics fail to capture non-Gaussian information, other summary statistics are desirable. Numerous summary statistics have been used to try to extract non-Gaussian information, including the wavelet scattering transform (e.g., Refs. [13–18]) and wavelet phase harmonics (e.g., Ref. [17]). Other approaches include higher-order N-point statistics (e.g., Refs. [19–23]), density-split clustering [24], topological descriptors such as Minkowski functionals [25–27], and machine-learning-based summaries like convolutional neural networks [28,29].

In this work, we compare the information content of selected summary statistics applied to the lensing of the cosmic microwave background and weak lensing fields using Fisher forecasts. Each summary statistic compresses the information in a field and typically depends on a handful of hyperparameters (such as the range of scales in a power spectrum or the number of different opening angles in three-point statistics). Hyperparameter choices change the effective scales and the dimensionality of the summary statistics;

*Contact author: kboone@g.harvard.edu

†Contact author: gvalogiannis@uchicago.edu

‡Contact author: mgatti@uchicago.edu

§Contact author: cdvorkin@g.harvard.edu

this can impact the stability of the covariance inversion in Gaussian-likelihood pipelines and, for simulation-based inference pipelines, inflate model complexity and training variance. Constraints become then highly dependent on hyperparameters. We address this by proposing and developing a novel learned binning approach, which produces compact and informative summaries.

In our approach, each summary statistic has access to all scales present in our maps. The statistics are then compressed via binning to the same size while maximizing the information content (the determinant of the Fisher matrix). By learning our binning and providing parameter constraints on distinct data sets, our method is robust to overfitting and yields a compressed statistic without sacrificing interpretability.

The rest of the paper is organized as follows: in Sec. II we introduce the simulations used in our analysis and explain how to use them to construct convergence maps. In Sec. III, we introduce the summary statistics used. In Sec. IV, we present an overview of our learned binning approach, which is then applied to simulated data in Sec. V. Finally, we conclude in Sec. VI. All parameter constraints are presented in Appendix A, and their convergence is tested in Appendix B. We test the convergence of our derivatives in Appendix C. Additional details on the learned binning approach are given in Appendix D. We verify the Gaussianity of our summary statistics required for a reliable Fisher forecast in Appendix E.

II. SIMULATED OBSERVABLES

We use weak gravitational lensing convergence as our simulated observable. Weak gravitational lensing distorts distant images and can be summarized by the lensing potential ψ , which is an integral of the gravitational potential Φ along the line of sight,

$$\psi(\vec{\theta}, D_S) = \frac{2}{c^2 D_S} \int_0^{D_S} dD_L \frac{D_{LS}}{D_L} \Phi(D_L \vec{\theta}, D_L), \quad (1)$$

where D_L , D_S , and D_{LS} are the angular distances from the observer to the lens, the observer to the source, and the lens to the source, respectively, c is the speed of light, and $\vec{\theta}$ is the angular position on the sky. From the lensing potential, we can then construct the convergence map κ as,

$$\kappa(\vec{\theta}, D_S) \equiv \frac{1}{2} \nabla_{\vec{\theta}}^2 \psi(\vec{\theta}, D_S). \quad (2)$$

In this work we exclusively use convergence maps integrated out to the cosmic microwave background (CMB) [$\kappa_{\text{CMB}}(\vec{\theta})$] and convergence maps integrated out to redshift ranges of galaxy surveys [$\kappa_{\text{WL}}(\vec{\theta})$]. These are defined in the following way:

$$\kappa_{\text{CMB}}(\vec{\theta}) \equiv \kappa(\vec{\theta}, D_{S,\text{CMB}}), \quad (3)$$

$$\kappa_{\text{WL}}(\vec{\theta}) \equiv \kappa(\vec{\theta}, D_{S,\text{WL}}), \quad (4)$$

where $D_{S,\text{CMB}}$ and $D_{S,\text{WL}}$ are the angular distances to the CMB and redshift bin of galaxies, respectively. We now turn to the simulations used to generate convergence maps.

A. Simulations: ULAGAM suite

The analysis in this work uses the ULAGAM suite of N-body simulations [30]. The suite consists of 2000 simulations at a fiducial cosmology, plus sets of 100 simulations where one parameter is varied at a time in a stepwise fashion for Fisher forecast purposes.

The simulations were run using PKDGRAV3 [31], and are performed in periodic cubic boxes of size $L = 1 h^{-1} \text{ Gpc} \approx 1.5 \text{ Gpc}$, initialized at redshift $z = 127$ and evolved with 512^3 dark matter particles. The initial conditions for all simulations were sourced from the Quijote suite [32]. The fiducial cosmological model is Λ cold dark matter (ΛCDM) and assumes the following parameters: matter density $\Omega_m = 0.317$, baryon density $\Omega_b = 0.049$, reduced Hubble constant $h = 0.671$, spectral tilt $n_s = 0.962$, and amplitude of matter density fluctuations at $8 h^{-1} \text{ Mpc}$ $\sigma_8 = 0.834$.

Our analysis focuses on forecasting constraints on Ω_m and σ_8 , the ΛCDM parameters determined by weak lensing. To evaluate the derivatives necessary for a Fisher forecast, we use simulations where Ω_m is shifted by ± 0.01 and where σ_8 is shifted by ± 0.015 . We verify in Appendix C that these parameter shifts are small enough for derivative estimates to converge. We use 1000 realizations for the fiducial cosmology, and 100 realizations each for variations in σ_8 and Ω_m . PKDGRAV3 automatically generates particle light cones in the form of HEALPix¹ pixel maps [33] at varying redshifts. These maps are provided at $N_{\text{SIDE}} = 1024$, corresponding to a pixel size of roughly 3.44 arcmin. The simulation box is tiled and repeated as needed to construct sufficiently large volumes for building full-sky light cones up to a given redshift. We convert the particle maps into overdensity maps and use them to generate weak-lensing convergence maps, which we discuss in the next subsection.

B. CMB lensing convergence

We follow the procedure described in Sec. III.1 of Omori [34] to create noiseless, full-sky κ_{CMB} maps for each of our simulations. In particular, we use the Born approximation to integrate the overdensity maps up to redshift $z = 3.5$. The CMB lensing power spectrum is then computed for the range $3.5 < z < 1089$, and a Gaussian realization based on this spectrum is generated and added to obtain the full κ_{CMB} map.

We generate mock survey κ_{CMB} maps by adding Gaussian noise and footprint masks to the full-sky, noiseless convergence maps, simulating various experiments. Specifically, we produce maps for *Planck*, ACT Data

¹<http://healpix.sourceforge.net>

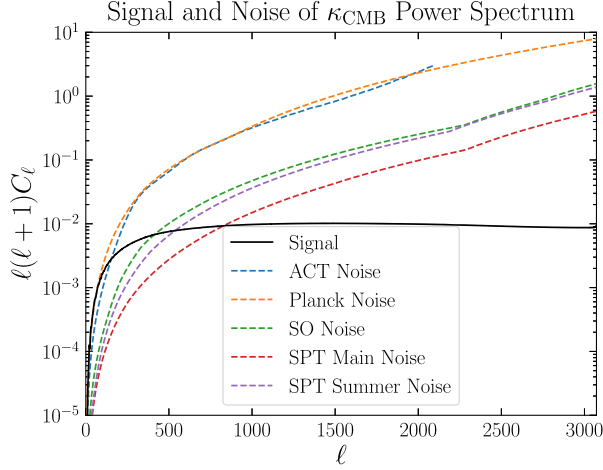


FIG. 1. We show the average noiseless auto power spectrum of κ_{CMB} (black) in comparison to noise power spectra from all CMB surveys used (dashed lines). For ACT, we only plot until $\ell = 2090$ as this is the limit of our noise power spectrum data. For SPT, we plot the noise power spectra from the main and summer fields separately. The cusps (for instance, at $\ell \approx 2200$ for SPT) occur because our noise power spectra are linear interpolations between specified points.

Release 6, SPT main and summer fields, and SO. The noise is modeled as purely Gaussian and generated from the corresponding noise power spectra, taken from Ade *et al.* [12] and Prabhu *et al.* [35]. For ACT, our noise power spectra only go to $\ell_{\text{max}} = 2090$, so we set all signals beyond this to zero. Signal and noise power spectra for κ_{CMB} are shown for all surveys used in Fig. 1.

For *Planck*, we further apply a Gaussian smoothing beam with a full width half maximum of 10 arcmin after adding noise. This mimics the instrument's beam. We ignore the beams of the other experiments because they are smaller than the pixel size of our maps. Survey masks are applied to each simulated map to match the respective sky coverage. For SPT, we consider the union of the main and summer fields, each region with its corresponding noise power spectrum applied. The coverage area of all surveys can be found in Table I. An example κ_{CMB} for *Planck* levels of noise is shown in Fig. 2.

We note that this procedure is idealized and produces simplified CMB lensing maps with Gaussian noise and no additional contaminants. These maps are sufficient for the purpose of this work, namely, a forecast-level assessment of the intrinsic non-Gaussian information in the lensing field, but they are not suitable for direct comparison with real data. In practice, CMB lensing is not measured directly but reconstructed from multifrequency temperature and polarization maps using, for example, quadratic estimators [36], after applying standard data-cleaning steps such as frequency filtering and survey-mask corrections. Since lensing is a small signal superimposed on the primary CMB anisotropies, the reconstructed maps are contaminated by the CMB itself and by reconstruction effects that must be

TABLE I. Coverage area for all CMB surveys considered in this work and their overlap area with *Euclid* Data Release 2.

Survey footprint areas (deg ²)		
Survey	Survey area	Overlap with <i>Euclid</i> Data Release 2
ACT	15 472	2903
<i>Planck</i>	24 915	7290
SPT	4001	2415
SO	27 229	3075

calibrated using simulations. Realistic CMB lensing mocks, therefore, require simulating unlensed CMB maps, lensing them with a full-sky convergence field, adding realistic anisotropic and non-Gaussian noise, and reconstructing the lensing field using the same estimators and corrections applied to the data [37]. The resulting maps naturally include non-Gaussian reconstruction noise and higher-order bias terms. Our aim here is more limited: to quantify, at the forecast level, the amount of intrinsic non-Gaussian information in the lensing field itself. For this purpose, modeling the noise as Gaussian is sufficient to capture the correct order-of-magnitude behavior, as we do not attempt to reproduce the detailed statistics of reconstructed CMB lensing maps.

C. Weak galaxy lensing convergence

To create mock κ_{WL} maps, we use a simplified version of the pipeline described in Gatti *et al.* [17], tailored to resemble *Euclid* Data Release 2 (DR2) data. First, we generate noiseless convergence maps as a function of redshift from the density field using the Born approximation.

We assume that the *Euclid* DR2 weak-lensing source sample is divided into six tomographic bins of roughly equal number density. For simplicity, we use only two tomographic bins: the lowest-redshift bin and the highest-redshift bin. The first (last) tomographic bin has a mean redshift of 0.33 (1.37) and a standard deviation of 0.09 (0.16). For each bin, we produce a convergence map by summing the redshift shells weighted by the corresponding redshift distribution. We add Gaussian shape noise, assuming an

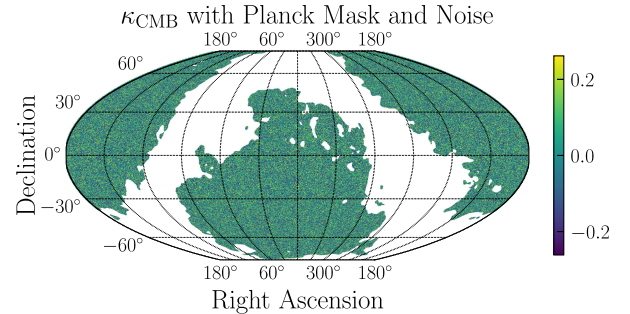


FIG. 2. We show κ_{CMB} with *Planck* levels of noise on *Planck*'s footprint.

ellipticity dispersion of $\sigma_e = 0.26$ and a galaxy number density of 30 galaxies per arcmin², uniformly distributed across the tomographic bins [3]. Finally, we mask the mock κ_{WL} maps by applying the *Euclid* DR2 footprint mask. For all $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ statistics, we apply both the *Euclid* and CMB-survey masks. The overlap coverage area for all surveys can be found in Table I.

III. SUMMARY STATISTICS

For κ_{CMB} maps, we quantify information using the angular power spectrum (C_ℓ) and the wavelet scattering transform (WST). For $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$, we generalize these statistics to the cross angular power spectrum and wavelet phase harmonics (WPH).

A. Power spectrum

For a function $A(\Omega)$ on the sphere, $a_{\ell m}^A$ values are defined as,

$$a_{\ell m}^A = \int A(\Omega) Y_{\ell, m}^*(\Omega) d\Omega, \quad (5)$$

with $Y_{\ell, m}$ denoting the spherical harmonics and Ω the solid angle. From the $a_{\ell m}$'s, we define the cross angular power spectra as,

$$C_\ell^{AB} = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} a_{\ell, m}^A a_{\ell, m}^{B*}. \quad (6)$$

The auto angular power spectra are defined with $A = B$. From here on, we will omit superscripts as the context will make it clear which power spectrum we are referring to. The C_ℓ 's are all computed on the masked fields described in Sec. II.

The C_ℓ 's are able to capture all Gaussian information about a field, but miss any information contained in non-Gaussianities. An example of a summary statistic for a single map that can capture non-Gaussian information is the WST, which we turn to next.

B. Wavelet scattering transform

The WST was initially proposed in Mallat [38] and then used in cosmology to reduce noise in simulations [39] and extract information from weak lensing maps [13] and galaxy clustering [14–16]. WST consists of convolving the initial map with a directional, multiscaled wavelet and taking the modulus of the result. Multiple successive convolutions can be performed to probe non-Gaussianities. An average over all pixels is used to convert any map into a scattering coefficient, the building block of this summary statistic.

More formally, starting with an input field, A_0 , one can construct,

$$A_{n+1} = |A_n \star \psi|. \quad (7)$$

In the above, ψ is the localized wavelet. In general, ψ will have a scale and a directional dependence, represented here by the integer-valued parameters j and l , respectively. In this case, one can obtain the first three orders of scattering coefficients as,

$$S_0 = \langle A_0 \rangle, \quad (8)$$

$$S_1(j_1, l_1) = \langle |A_0 \star \psi_{j_1, l_1}| \rangle, \quad (9)$$

$$S_2(j_1, l_1, j_2, l_2) = \langle ||A_0 \star \psi_{j_1, l_1} \star \psi_{j_2, l_2}| \rangle, \quad (10)$$

where $\langle \dots \rangle$ is a spatial average. As in other works (e.g., Refs. [17, 18]), we only use S_1 and S_2 coefficients. If the overlap of wavelets $\psi_{j, l}$ and $\psi_{j, l+1}$ is small and we have statistical isotropy, we can average over orientations without losing information: $S_1(j_1, l_1) \rightarrow S_1(j_1)$ and $S_2(j_1, l_1, j_2, l_2) \rightarrow S_2(j_1, j_2, |l_1 - l_2|)$. In this work, we assume isotropy for all relevant summary statistics and average over opening angles. Additionally, not all scale combinations contribute to S_2 . As shown in Cheng and Ménard [40], only S_2 coefficients with $j_2 > j_1$ are significant.

In this work, we use the Morlet wavelets, given by,

$$\psi_{j, l}(\vec{x}) = \frac{1}{\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) [\exp(i\vec{k}_0 \cdot \vec{x}) - \beta], \quad (11)$$

$$\tilde{\psi}_{j, l}(\vec{k}) = \exp\left(-\frac{(\vec{k} - \vec{k}_0)^2 \sigma^2}{2}\right) - \beta \exp\left(-\frac{k^2 \sigma^2}{2}\right), \quad (12)$$

where $\sigma = 0.8 \times 2^j$, $k_0 = 0.75\pi \times 2^{-j}$, $\arg(\vec{k}_0) = \theta$, with $\theta = (\frac{L}{2} - 1 - l)\pi/L$. The counterterm $\beta = \exp(-\frac{k_0^2 \sigma^2}{2})$ is introduced so that $\tilde{\psi}(0) = 0$. The j index will take integer values between 0 and $J - 1$, where J controls the number of scales and is at most $\log_2(\text{Image Size}) - 1$. This keeps the wavelet peaks in Fourier space roughly between the fundamental frequency and the Nyquist frequency of the image. The l values range between 0 and $L - 1$, where L specifies the number of different angles to probe. While Morlet wavelets traditionally have the above dyadic spacing, we also experiment with finer spacing. Details on this will be given explicitly before any results are shown.

To apply the wavelet scattering transform, the input image must be two-dimensional and flat. To achieve this with spherical data, we split the data into 768 disjoint patches, which are the HEALPix pixels at $NSIDE = 8$. The center of each of these patches is rotated to the north pole of the sphere. From here, gnomonic projection is used to project the high-resolution patch onto a 256×256 grid [as in [17]], which sets our maximum J value at $J = 7$. Due to spherical distortions, not all HEALPix pixels are mapped one-to-one onto a pixel in the final grid. We do not attempt to correct for this effect in this work. After obtaining

scattering coefficients for each patch, we average them to obtain scattering coefficients for the entire map.

Since our data is masked, we use the convention to only use a patch if at least 20% of its HEALPix pixels are unmasked. The unmasked fraction for a patch is used to weight all of its scattering coefficients prior to averaging across patches. We use the wavelet scattering repository² for calculating scattering coefficients after modifying it to allow for nondyadic spacing.

C. Wavelet phase harmonics

WPH is a generalization of some of the concepts of WST to two maps, which we will denote as A and B . Since WPH uses two maps instead of the one used for WST, it allows for the study of cross-correlations. Unlike WST, each map is only ever convolved with one wavelet, which we denote as,

$$A_{j,l} = A \star \psi_{j,l}. \quad (13)$$

As in WST, $\psi_{j,l}$ is a localized wavelet characterized by a scale parameter j and a directional parameter l . In general, $A_{j,l}$ will be complex valued. We apply a phase harmonic to this, defined as,

$$\text{PH}(r \exp(i\theta), q) = r \exp(iq\theta), \quad (14)$$

with r , θ , and q real-valued numbers.

From the equation above, we can define WPH coefficients in general. As in WST, because we are assuming isotropy, the only angular dependence will be in $\Delta l \equiv |l_2 - l_1|$, and we can write our WPH coefficients as,

$$\begin{aligned} \text{WPH}_{q_1 q_2}(A, B, j_1, j_2, \Delta l) \\ = \langle \text{PH}(A_{j_1, l_1}, q_1) \text{PH}(B_{j_2, l_2 + \Delta l}, q_2) \rangle. \end{aligned} \quad (15)$$

The WPH coefficients that we use are,

$$S_{00}(A, B, j) = \langle |A_{j,l}| |B_{j,l}| \rangle, \quad (16)$$

$$S_{01}(A, B, j) = \langle |A_{j,l}| |B_{j,l}| \rangle, \quad (17)$$

$$S_{10}(A, B, j) = \langle |A_{j,l}| |B_{j,l}| \rangle, \quad (18)$$

$$S_{11}(A, B, j) = \langle A_{j,l} B_{j,l} \rangle, \quad (19)$$

$$C_{01}(A, B, j_1, j_2, \Delta l) = \langle |A_{j_1, l_1}| |B_{j_2, l_2 + \Delta l}| \rangle, \quad (20)$$

$$C_{10}(A, B, j_1, j_2, \Delta l) = \langle A_{j_1, l_1} |B_{j_2, l_2 + \Delta l}| \rangle, \quad (21)$$

with $j_1 < j_2$ for C_{01} and $j_2 < j_1$ for C_{10} .

These statistics are chosen following Gatti *et al.* [17], with the only change being that we allow for more opening

angles in C_{01} and C_{10} statistics. For simplicity, we only consider the real component of the final statistics.

In this work, we use the package PyWPH [41] to apply WPH and use the bump-steerable wavelets, which are given by the following mother wavelet:

$$\psi(\vec{k}) = 0.693 \exp\left(\frac{-(k - \xi_0)^2}{\xi_0^2 - (k - \xi_0)^2}\right) \cos^3(\arg(\vec{k})). \quad (22)$$

The equation above is for Fourier wave numbers in the range $0 < k < 2\xi_0$ and $|\arg(k)| < \pi/2$; for all other values $\psi = 0$. The normalization and the central frequency $\xi_0 = 0.85\pi$ are set as in Regalado-Saint Blancard *et al.* [41]. All other wavelets are generated by dilating or rotating the mother wavelet. In real space,

$$\psi_{j,l}(\vec{x}) = 2^{-j} \psi(2^{-j} \text{Rot}_{-\pi l/L}(\vec{x})). \quad (23)$$

In the above, “Rot” means to rotate the vector by the subscripted angle. We probe $L = 4$ different angles.

To apply WPH, the input images must be two-dimensional and flat, as for WST. We follow the same procedure as we used for WST to project patches of the sphere onto 256×256 grids. As in the case of WST, we only use a patch if at least 20% of its HEALPix pixels are unmasked. The unmasked fraction for a patch is used to weight the WPH coefficients from each patch before averaging across patches to obtain the WPH coefficients for the entire map.

IV. ANALYSIS

We use the Fisher information approach to compare the constraining power of our different summary statistics. In doing so, we assume that our summary statistics are Gaussian distributed, which we test and confirm in Appendix E. To fairly compare different summary statistics, we compress them with a novel learned binning approach. This approach compresses the statistics to a set number of bins while maximizing the determinant of the Fisher matrix.

A. Fisher forecast

To generate our covariance matrix C , we use 1000 fiducial simulations. Derivatives for the Fisher forecast are estimated numerically using 100 realizations per parameter shift. We concatenate our derivative vectors into a matrix D , so we can write,

$$F = D^T C^{-1} D. \quad (24)$$

To unbiased our inverse covariance, we apply the Hartlap correction [42]; however, this correction alone does not yield an unbiased Fisher matrix. Noise in derivative estimates causes the Fisher matrix to systematically overestimate the constraining power of a summary statistic [43].

²https://github.com/SihaoCheng/scattering_transform

To address this, Coulton and Wandelt [43] introduce a compressed Fisher matrix, F^{comp} . This Fisher matrix is constructed as in Eq. (24), but using summary statistics compressed with the MOPED [44] algorithm. Perfect compression of the summary statistics would lead to no information loss, but the compression relies on the matrices D and C . Coulton and Wandelt [43] show that F^{comp} will systematically underestimate the constraining power if D and C are estimated from simulations via Monte Carlo methods before compressing the statistics. Taking the geometric mean of F and F^{comp} gives the combined Fisher matrix, F^{Combined} , in which the opposing biases of the two inputs cancel, yielding unbiased constraints. It is this combined Fisher matrix and the resulting parameter covariance matrix, $(F^{\text{Combined}})^{-1}$, which we use for all shown constraints.

The construction of F^{comp} requires a random splitting of the derivatives, leading to stochasticity in F^{Combined} . To reduce shot noise, we generate 100,000 realizations of the parameter covariance matrix with different random splits of the derivatives. Our final parameter covariance matrix is the average of the middle 90,000 realizations ordered by determinant. We cut off the most extreme 5,000 realizations on either end to reduce the impact of extreme outliers, which can occur in this method. This averaging reduces shot noise while still leaving us sensitive to systematic convergence issues, which we test for.

Before constructing the combined Fisher matrix, we need to compress our summary statistics to obtain an invertible data covariance matrix. Compression is necessary since the finite number of realizations used to evaluate the covariance matrix places an upper bound on its rank. Since all of our summary statistics have scale dependence, in this work we implement a novel learned binning (over scale) approach to compress our statistics, while retaining interpretability. We explain our method below.

B. Learned binning compression

In this section, the initial summary statistic will be M dimensional, and it will be compressed to N dimensions, with $N < M$. Any binning scheme can be viewed as a particular linear transformation within $\mathbb{R}^{N \times M}$. We will refer to the general binning matrix as B . To compress a data vector $\vec{d}$, we perform $B\vec{d}$.

Initially, with $D \in \mathbb{R}^{M \times 2}$ (we have two cosmological parameters) and $C \in \mathbb{R}^{M \times M}$, one would be unable to calculate the Fisher matrix as in Eq. (24) if M exceeds the number of simulations. With a binning matrix, one would instead perform,

$$F = (BD)^T (BCB^T)^{-1} BD. \quad (25)$$

In this way, we are now inverting an $N \times N$ matrix instead of an $M \times M$ matrix. Therefore, we want the value of N to be much smaller than the number of fiducial simulations. Everything so far simply applies to general $N \times M$ linear

transformations beyond just binning matrices. Compared with general linear transformations, binning is more intuitive and interpretable while having fewer free parameters, thereby reducing the risk of overfitting.

To enable low-dimensional parametrizations of binning matrices, we use a function $f(x): [1, M] \rightarrow [1, N]$ to specify the positions of nonzero matrix entries. Specifically, the m^{th} column of the binning matrix has a single nonzero entry equal to 1 at the integer index closest to $f(m)$. In this work, $f(x)$ is a linear interpolation between $(1,1)$, learned pivots, and (M,N) . Further details on the conditions that $f(x)$ must satisfy can be found in Appendix D. An example of such a construction is shown in Fig. 3. We will often refer to the summary statistic index and bin number of the pivot as the “ x ” or “ y ” components of the pivot.

To compress summary statistics, we first pick a number of pivots for the binning matrix. Some summary statistics (cross- C_ℓ ’s, WST, and WPH, in this work) can be thought of as consisting of multiple subclasses of statistics that depend exclusively on scale. Taking WST as an example, S_1 and S_2 coefficients cannot be naturally compared based on scale. Additionally, within the S_2 coefficients, statistics with different opening angles cannot be naturally ordered based on scale. Therefore, for WST, we split the coefficients into five subclasses that have only scale dependence (S_1 and the four different opening angle components of S_2). These statistics are ordered individually by scale and then appended to form our full ordered summary statistic. We fix the x coordinates of certain pivots to the transitions between subclasses. This effectively introduces parameters which, in combination with the total number of bins, exclusively control how many bins are allocated to each subclass.

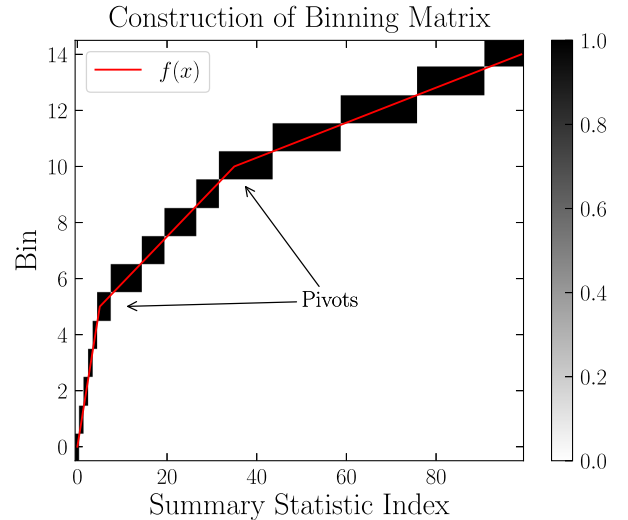


FIG. 3. Construction of a binning matrix from a function $f(x)$. Our function is a linear interpolation between labeled pivots. This binning matrix compresses a summary statistic of length 100 to 15 bins.

Pivots are therefore defined by two free parameters if their x coordinate is not specified and one free parameter otherwise. If there are p pivots in total and p_x of them have specified x coordinates, the pivots can collectively be parametrized by a point in a space with dimension $2(p - p_x) + p_x$. We call this the “pivot space.”

Each point in pivot space corresponds to a binning of the summary statistics that can be used to create a unique Fisher matrix. To retain as much information as possible, we attempt to maximize $\det(F)$. For the sake of efficiency, this optimization uses the Fisher matrix from Eq. (24) instead of the combined Fisher matrix. The relatively low number of free parameters ($\mathcal{O}(\sim 10)$ in this work) reduces the probability of overfitting compared to a general linear compression with $N \times M$ free parameters. An outline of our optimization algorithm is given in Appendix D.

When only using one pivot, the pivot space is two-dimensional, and $\det(F)$ can be plotted as a heatmap. An example of this for *Planck* κ_{CMB} C_ℓ ’s is shown in Fig. 4. This plot shows that $\det(F)$ has a genuine signal as a function over pivot space, but that there is also numerical noise that comes from having a finite number of simulations. Since we maximize $\det(F)$, we could end up fitting to a local maximum in the noise, which would lead to us overestimating the strength of constraints.

To avoid overfitting, we randomly split our fiducial and derivative simulations in half, learn pivot positions on one half, and then generate contours from the other half. While we use the Fisher matrix from Eq. (24) to learn how to bin our summary statistics, once compressed, we use the combined Fisher matrix to create contour plots. We repeat this process for 1000 different splits and get a covariance matrix for our cosmological parameters for

each split. We take our final covariance matrix to be the average of these matrices.

We use 15 bins since we find that this choice yields binned summary statistics that follow a Gaussian distribution (see Appendix E). With 15 bins, it would be feasible to determine the size of each bin individually. However, this approach would struggle to generalize to future works if more bins were used. Determining the optimal placements of bin boundaries for N bins consists of maximizing $\det(F)$ over an $N - 1$ dimensional space. When using our pivot approach, the dimension of the pivot space, $\det(F)$, is maximized over and is completely decoupled from N . This allows one to control the number of free parameters.

In general, we note that compressing summary statistics is a nontrivial problem, and a variety of methods have been proposed in the literature [45]. Many of these techniques, however, require training on simulations at varying cosmologies, which is not feasible in our setting, where only a fiducial covariance matrix and its derivatives are available. Other alternatives in this regime include principal component analysis, finding optimal subsets of observables [46], and the widely used MOPED algorithm [44]. In principle, MOPED would be the most principled choice, as it is designed to maximize the Fisher information. In practice, however, its performance is limited in our case: it relies on stable estimates of derivatives and on inverting a high-dimensional covariance matrix, both of which can be noisy when only a modest number of simulations are available. This issue is exacerbated by the need to split the simulations into independent subsets for training and validation to avoid overfitting, which further reduces the effective number available. Although our learned-binning scheme also maximizes Fisher information, the compression matrix is constrained to a simple, nearly diagonal form that depends on only a handful of pivot parameters. This structure makes it substantially less sensitive to noise than the dense compression vectors produced by MOPED, while preserving interpretability. Using cross-validation further mitigates overfitting to fluctuations in the Fisher matrix, yielding a robust and practical method.

V. RESULTS

We now turn to forecasted constraints on Ω_m and σ_8 after binning our summary statistics. We first compare auto C_ℓ ’s and WST applied to κ_{CMB} maps in Sec. VA. Next, we look at cross- C_ℓ ’s and WPH applied to $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ maps in Sec. VB. We note that results for WST and WPH in the next section are not combined with those from the power spectrum.

A. CMB lensing

For CMB lensing convergence maps, we show results for the angular power spectrum and also for WST, as our summary statistics.

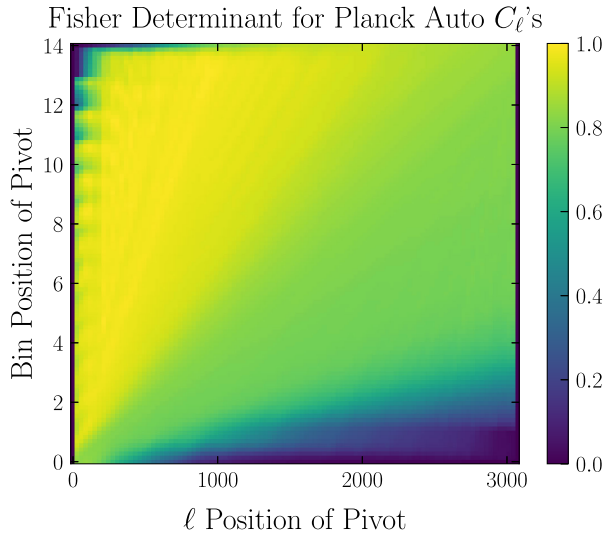


FIG. 4. Determinant (normalized to have a maximum of 1) of the Fisher matrix from Eq. (24) over pivot space for *Planck* κ_{CMB} C_ℓ ’s. Only one pivot is used to enable visualization via a heatmap, so the axes denote the ℓ and bin positions of the pivot.

For WST, we do not use dyadic sequencing; instead, we have j from Eq. (11) take 50 different values from 0 to 6. This allows greater flexibility in the learned binning by increasing the size of the initial data vector. For angular spacing, we pick $L = 4$. With these choices, there are 5150 WST coefficients. For C_ℓ 's, our resolution limits the maximum ℓ to 3071, so after discarding the dipole, we have 3070 C_ℓ values.

For the learned binning, ordering the original summary statistic is important. Entries that are near each other will likely be binned together, so we want these entries to be highly correlated to minimize information loss. For C_ℓ 's, this is done by ordering based on ℓ . S_1 coefficients are likewise ordered based on j . For S_2 coefficients, we have two scales and an opening angle, making the ordering scheme more arbitrary. In this work, we divide the S_2 coefficients into four different subclasses based on the opening angle. In each subclass, we order based on j_2 . When multiple coefficients have the same opening angle and j_2 value, we order based on j_1 . Across surveys, we find that when looking at pivot positions, the S_1 coefficients are the most heavily weighted.

For the C_ℓ 's, we find that allowing one pivot can slightly improve the constraints, but allowing two yields little additional gain. Therefore, for C_ℓ 's we search for one pivot. For WST, we lock four pivots to the transitions from S_1 to S_2 and between the different opening angles of S_2 . Since one dimension of the pivot is locked, this has four degrees of freedom. We find improvements by adding one

TABLE II. 1σ constraints on Ω_m and σ_8 for C_ℓ 's and WST of κ_{CMB} maps for all surveys considered in this work.

κ_{CMB} constraints: $\sigma(\Omega_m)[\times 100]/\sigma(\sigma_8)[\times 100]$		
Survey	C_ℓ 's	WST
ACT	1.43/0.72	1.54/0.77
<i>Planck</i>	1.51/0.73	1.56/0.73
SPT	0.90/0.59	1.00/0.66
SO	0.52/0.35	0.59/0.40

additional arbitrary pivot that can be placed anywhere, resulting in a six-dimensional search space for the pivots, with no significant gains beyond this. Across surveys, we find that the first ~ 1000 C_ℓ 's are consistently the most heavily weighted.

After finding the pivots, we use the combined Fisher forecast to get constraints on the cosmological parameters Ω_m and σ_8 . For all of our surveys, we find that WST and C_ℓ 's give the same constraints on both Ω_m and σ_8 to within 15%, showing that there is not much non-Gaussian information to extract from κ_{CMB} . *Planck* constraints are shown in Fig. 5, and constraints for all surveys are shown in Appendix A. Additionally, all 1σ constraints can be found in Table II.

When we look at the covariance of the two binned statistics, we find the expected positive correlation between C_ℓ 's and S_1 coefficients. Additionally, the covariance matrix of the WST coefficients has significant off-diagonal entries, primarily between the S_2 coefficients. This highlights that while our learned binning algorithm can be useful in diagonalizing the covariance matrix, it is only possible if correlations are originally among entries close enough together to be in the same bin. This corresponds to an initial nonbinned covariance matrix with significant entries only near the diagonal.

B. CMB lensing $\times$ weak lensing

For the cross-correlation of the convergence of WL and CMBL fields ($\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$), we use cross- C_ℓ 's and WPH as our summary statistics. For WPH we use $J = 7$ and $L = 4$, which leads to 392 coefficients. We use dyadic sequencing for WPH (unlike WST) to reduce runtime. Generating WPH coefficients for one map and one survey takes about 10 min in our implementation.

For cross- C_ℓ 's, we have two spectra coming from our two tomographic bins for WL, leading to 6140 C_ℓ 's. We order each power spectrum based on ℓ and then append them. We lock a pivot at the transition between tomographic bins and add two more arbitrary pivots so that each tomographic bin can get one pivot like the auto C_ℓ 's of the previous section. This leads to a five-dimensional search space for the pivots, and we find no significant improvements beyond this. Across surveys, we find that the first ~ 500 C_ℓ 's from the higher redshift tomographic bin are the most heavily weighted.

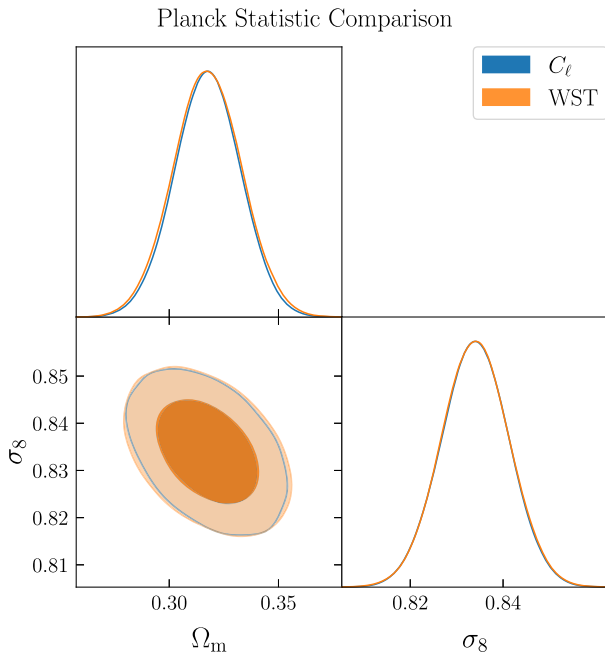


FIG. 5. We show 1σ and 2σ constraints on Ω_m and σ_8 from *Planck* κ_{CMB} , using the C_ℓ 's (blue) and the WST coefficients (orange). These constraints are obtained using our learned-binning approach. The constraints are the same to within 5%.

For WPH, for each tomographic bin we have S_{00} , S_{01} , S_{10} , S_{11} , C_{01} , and C_{10} coefficients. For C_{01} and C_{10} , we also have four different opening angles that we split into different subclasses of statistics. From this, for each of the two tomographic bins, we have 12 subclasses within the WPH summary statistic, leading to a total of 24 subclasses. Within each subclass, we order based on j . For C_{01} (C_{10}), we order based on j_2 (j_1). When multiple coefficients have the same opening angle and j_2 (j_1) value, we order based on j_1 (j_2). We lock pivots at the transitions between each subclass of the WPH statistics. Since this configuration already has a relatively large number of degrees of freedom for binning, we do not add any additional pivots. Across surveys, S_{00} and S_{11} consistently receive the largest number of bins.

As an example of the power of the learned binning, we can compare constraints with the learned binning compared to an initial binning. As the simplest possible initial binning to compare to, we look at linear binning with no pivots. Constraints are shown in Fig. 6 for ACT WPH. Constraints improve by factors of 2.7 for Ω_m and 2.5 for σ_8 when using a learned binning. We verify the convergence of this method in Appendix B.

We find that constraints on Ω_m and σ_8 are both significantly tighter when using WPH compared to cross- C_ℓ 's. Constraints on Ω_m improve by factors between 2.3 for *Planck* $\times$ *Euclid* DR2 and 3.4 for SPT $\times$ *Euclid* DR2. Improvements for σ_8 are between factors of 2.2 for *Planck* $\times$ *Euclid* DR2 and 3.4 for SPT $\times$ *Euclid* DR2. In Fig. 7 we show this result for *Planck* $\times$ *Euclid* DR2.

Additionally, all 1σ constraints can be found in Table III. Full contour results for other surveys are provided in Appendix A. All 1σ constraints when no pivots are learned are also presented in Table IV. Cross- C_ℓ 's see no significant changes in constraining power when using learned binning, with constraints fluctuating by under 10%. For our CMB lensing auto statistics, C_ℓ 's also see constraint changes of under 10% when using learned binning. Larger impacts are observed with WST, where constraints on Ω_m are up to ~ 1.5 times tighter in ACT. Results do not change appreciably for the power spectrum statistics if logarithmic binning is used instead of linear binning.

Our results show that WPH can extract information from non-Gaussianities in the lensing. Compared to the lack of gains found for CMB lensing alone, this is expected: weak-lensing–CMB-lensing cross-correlations peak at significantly lower redshift ($z \sim 1$) compared to CMB lensing autocorrelations ($z \sim 2$ –6). At these lower redshifts, the matter distribution is more evolved and inherently more non-Gaussian, providing WPH with additional information to extract beyond that available to power spectra alone. These findings are consistent with those of previous WL applications in the literature (e.g., Refs. [13,17,18,26,29]).

To test the convergence of our results, we repeat this entire process (including finding pivots) while lowering the total number of derivatives used. We examine how the constraints vary with the number of derivative realizations used and find that the results have sufficiently converged. All results are shown in Appendix B.

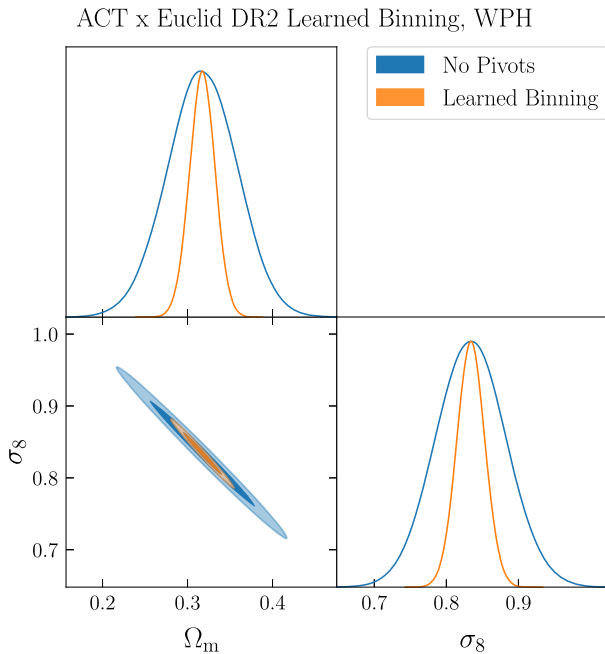


FIG. 6. Constraints for ACT $\times$ Euclid DR2 ($\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$) from WPH. The learned binning (orange) outperforms the initial binning with no pivots (blue). Constraints are improved by factors of 2.7 for Ω_m and 2.5 for σ_8 .

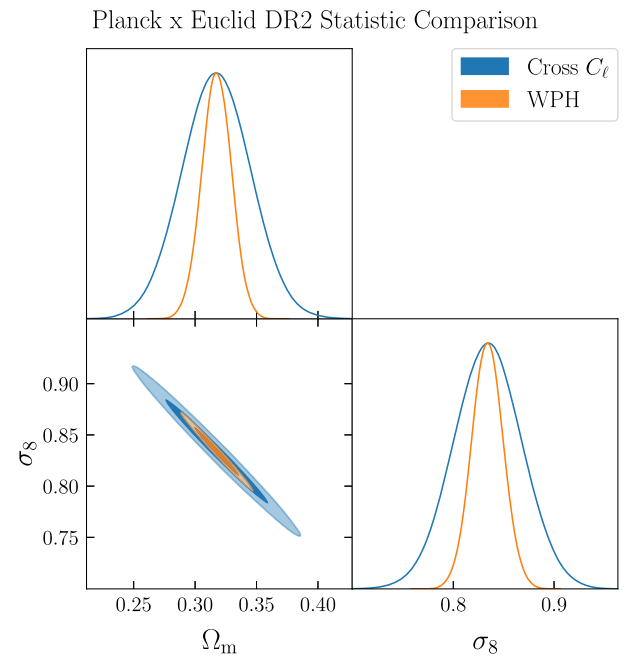


FIG. 7. Constraints for *Planck* $\times$ *Euclid* DR2 ($\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$) from cross- C_ℓ 's and WPH using our learned-binning approach. The Ω_m and σ_8 constraints are improved by factors of 2.3 and 2.2, respectively, when going from cross- C_ℓ 's to WPH.

TABLE III. 1σ constraints on Ω_m and σ_8 for cross- C_ℓ 's and WPH of $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ maps for all surveys.

$\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ constraints: $\sigma(\Omega_m)[\times 100]/\sigma(\sigma_8)[\times 100]$		
Survey	Cross- C_ℓ 's	WPH
ACT $\times$ Euclid DR2	4.09/4.97	1.52/1.96
Planck $\times$ Euclid DR2	2.79/3.39	1.20/1.56
SPT $\times$ Euclid DR2	2.57/3.29	0.76/0.95
SO $\times$ Euclid DR2	2.67/3.39	0.86/1.10

TABLE IV. 1σ constraints on Ω_m and σ_8 for cross- C_ℓ 's and WPH of $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ maps for all surveys. For these results, no pivots have been learned before making the constraints.

$\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ no learned pivots constraints: $\sigma(\Omega_m)[\times 100]/\sigma(\sigma_8)[\times 100]$		
Survey	Cross- C_ℓ 's	WPH
ACT $\times$ Euclid DR2	3.88/4.86	4.07/4.84
Planck $\times$ Euclid DR2	2.89/3.51	2.80/3.30
SPT $\times$ Euclid DR2	2.41/3.08	1.68/1.96
SO $\times$ Euclid DR2	2.56/3.24	1.80/2.16

VI. CONCLUSION

In this work, we study CMB lensing convergence and galaxy weak-lensing convergence data using a variety of summary statistics. To reduce the dimensionality of our summary statistics, we introduce a novel, learned-binning approach. Like other data reduction methods such as MOPED [44], our algorithm attempts to maximize Fisher information in the compressed statistic. Since our method uses fewer parameters, it is less susceptible to noise from a limited number of simulations. Our method also never requires inverting the original data covariance matrix, which becomes impossible if the number of simulations is less than the length of the summary statistic. By construction, our learned binning approach is robust to overfitting and can improve constraints relative to linear binning by up to a factor of 2.7 for the surveys and observables considered in this work.

For CMB lensing convergence data, we compare the higher-order statistics of the wavelet scattering transform to the standard angular power spectrum. We find that these summary statistics lead to similar constraints for all of the surveys considered. When we cross our CMB lensing simulated data with Euclid DR2 galaxy weak lensing convergence, we find that the higher-order statistic of the wavelet phase harmonics significantly outperforms the cross power spectrum. Constraints are improved by factors of up to 3.4. Compared to constraints from cosmic microwave background lensing convergence data alone, the cross-survey constraints exhibit stronger degeneracies. All contour plots are presented in Appendix A.

With WPH showing the most promise of the two higher-order statistics, further generalizations of WPH could prove useful. One potential next step would be investigating the use of WPH statistics besides S_{00} , S_{01} , S_{10} , S_{11} , C_{01} , and C_{10} . WPH and WST also require us to break the sky into patches, which limits the scales we can probe. Generalizing our pipeline to spherical wavelets [47] could circumvent this issue. Spherical wavelets have been previously used for CMB analysis [48,49].

This work demonstrates that, even though higher-order statistics are largely unnecessary for CMB lensing applications, they can substantially improve cosmological parameter constraints from galaxy weak-lensing–CMB-lensing cross-correlation data.

We also provide a novel method for comparing multiple summary statistics without sacrificing interpretability, which will become increasingly important as new summary statistics are introduced.

A physical interpretation of these results follows from the discussion in the previous section. WL-CMBL cross-correlations peak at lower redshifts ($z \sim 1$) than CMBL convergence autocorrelations ($z \sim 2-6$), where the matter distribution is more evolved and therefore more non-Gaussian. This makes non-Gaussian statistics particularly powerful in the cross-correlation case. While applications of non-Gaussian statistics to weak lensing alone are already well established (e.g., Refs. [13,17,18,22,23,26,29]), their use in WL-CMBL convergence cross-correlation data has not been explored before. One reason is that it requires forward modeling of the CMB lensing reconstruction process, including all relevant observational systematics, which is highly nontrivial. Nonetheless, the results presented here suggest that such an effort would be worthwhile.

ACKNOWLEDGMENTS

The authors would like to thank Anmol Raina for his contributions during the initial stages of this project and for insightful discussions. G. V. acknowledges the support of the Eric and Wendy Schmidt AI in Science Fellowship at the University of Chicago, a program of Schmidt Sciences. M. G. is supported by the Kavli Institute for Cosmological Physics at the University of Chicago through an endowment from the Kavli Foundation. This work was supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, [50]).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: CONSTRAINTS FOR ALL SURVEYS

In this appendix, we show all the parameter constraints studied in this work. All the constraints shown here are obtained using our learned-binning approach. In Fig. 8 we show constraints for each summary statistic and survey considered in this work. For each survey we also compare power spectrum and wavelet constraints directly for κ_{CMB} statistics in Fig. 9 and for $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ statistics in Fig. 10.

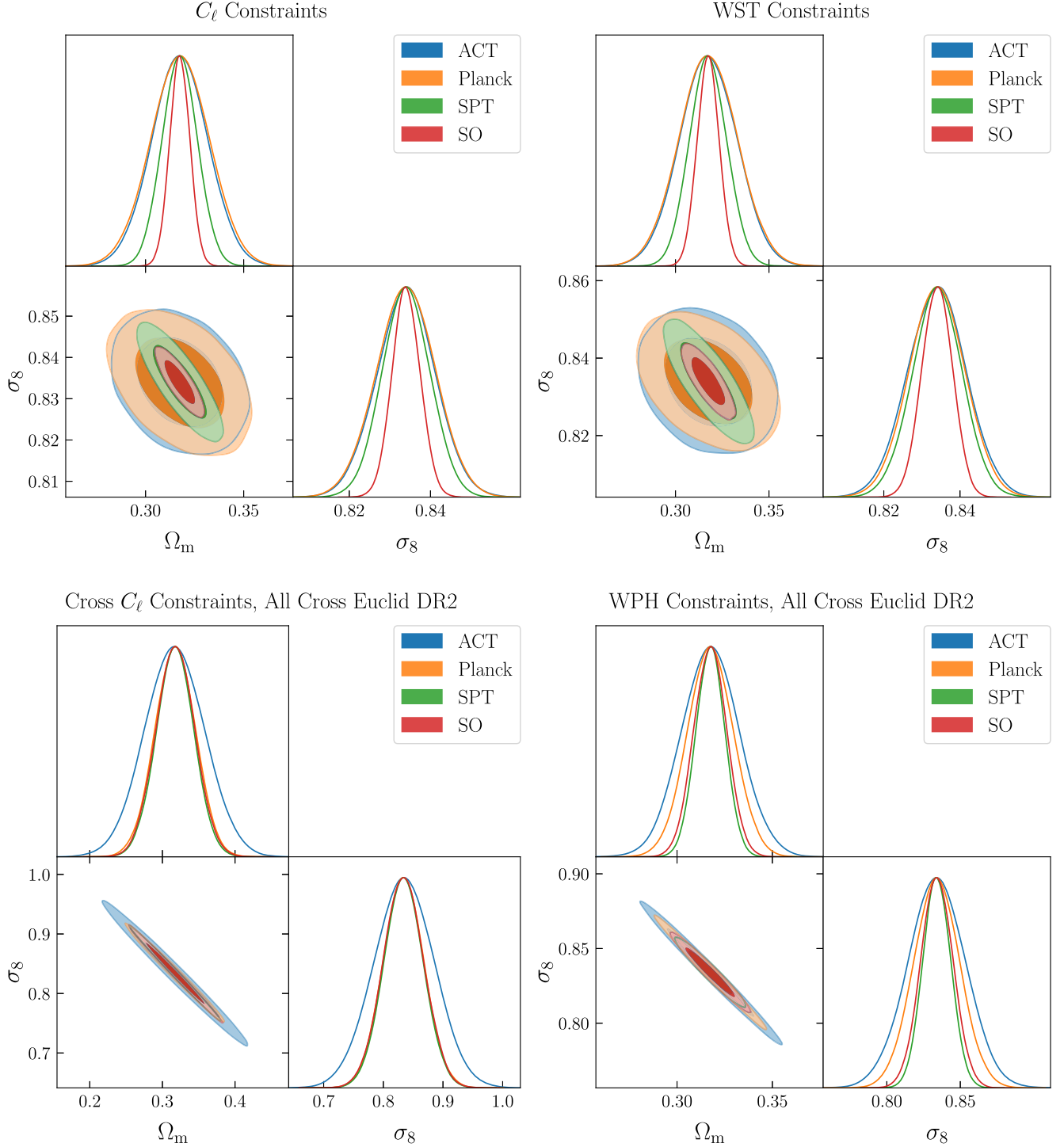
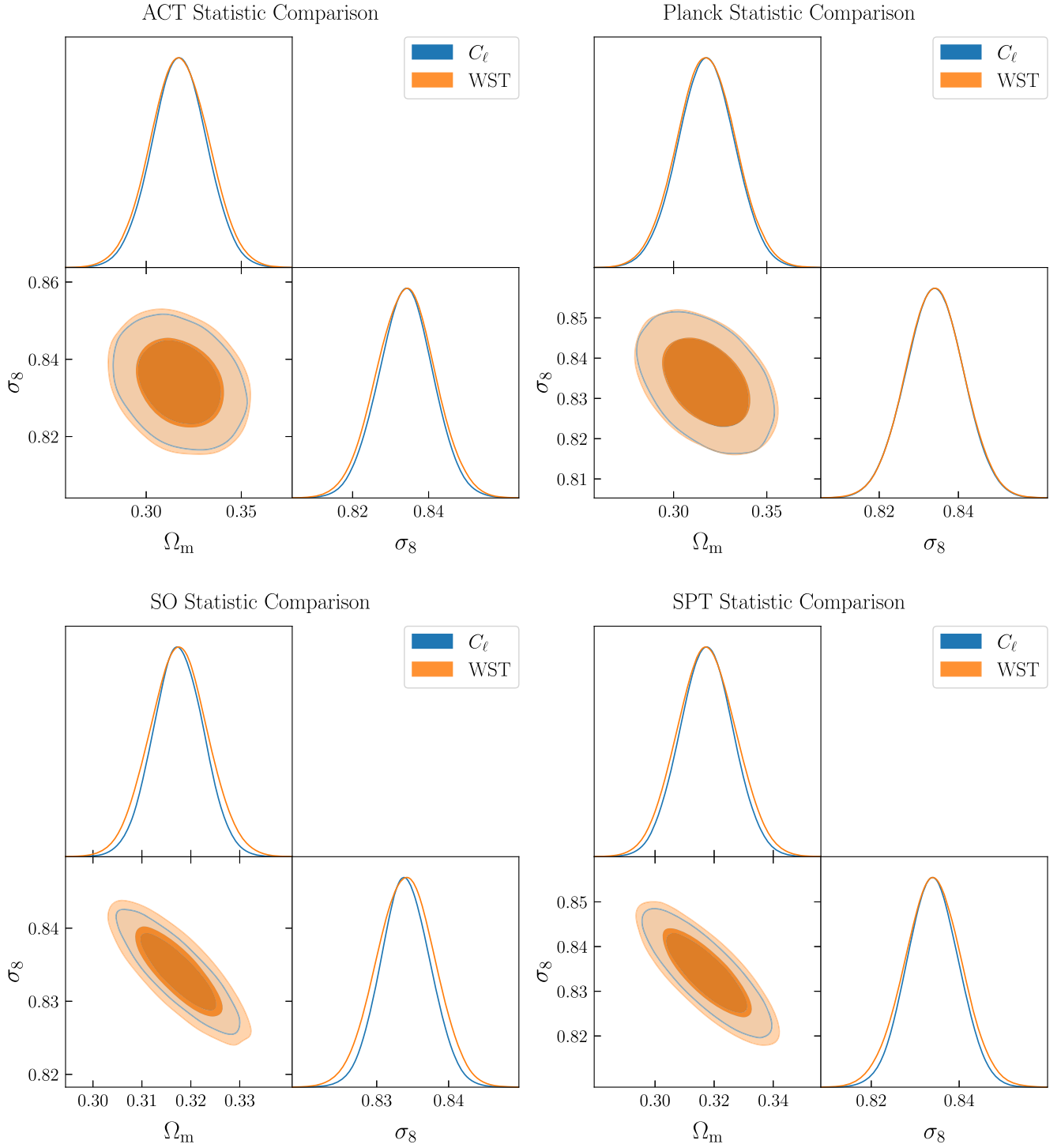


FIG. 8. Constraint comparisons between surveys for all summary statistics used. The upper plots use κ_{CMB} while the bottom plots use $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$.

FIG. 9. Constraint comparisons between C_ℓ 's and WST for κ_{CMB} maps.

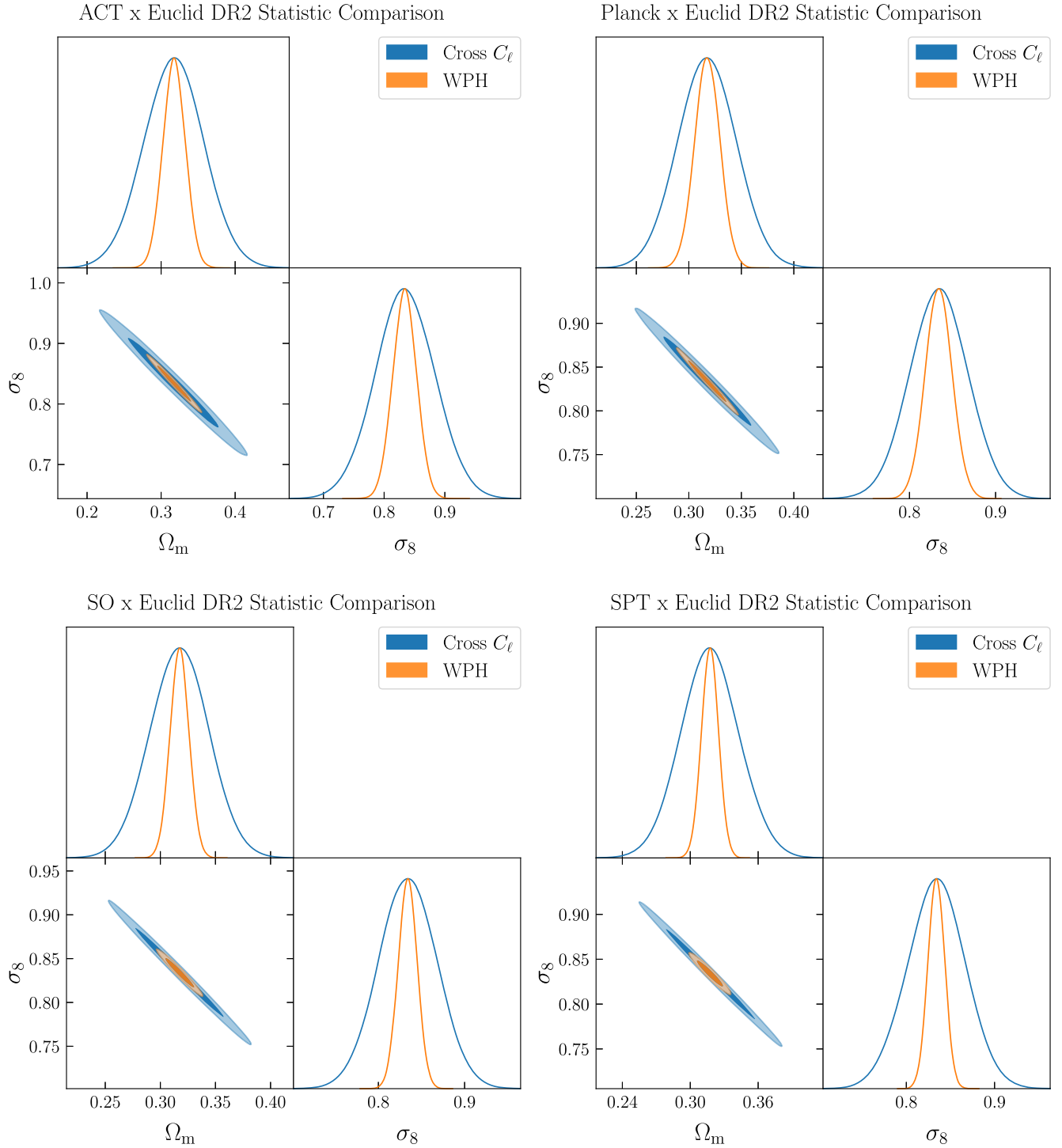


FIG. 10. Constraint comparisons between cross- C_ℓ 's and WPH for $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$.

APPENDIX B: CONTOUR CONVERGENCE FOR ALL SURVEYS

We test the convergence of our methodology by reducing the number of available derivatives and observing how the contours change. The number of realizations for generating both Ω_m and σ_8 derivatives is changed simultaneously.

For brevity, we only show the convergence contours for $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$ statistics in Figs. 11 and 12. Constraints for auto κ_{CMB} statistics all converge to similar levels. Additionally, we show how 1σ constraints on Ω_m and σ_8 converge individually for cross- C_ℓ 's and WPH, in Figs. 13 and 14, respectively.

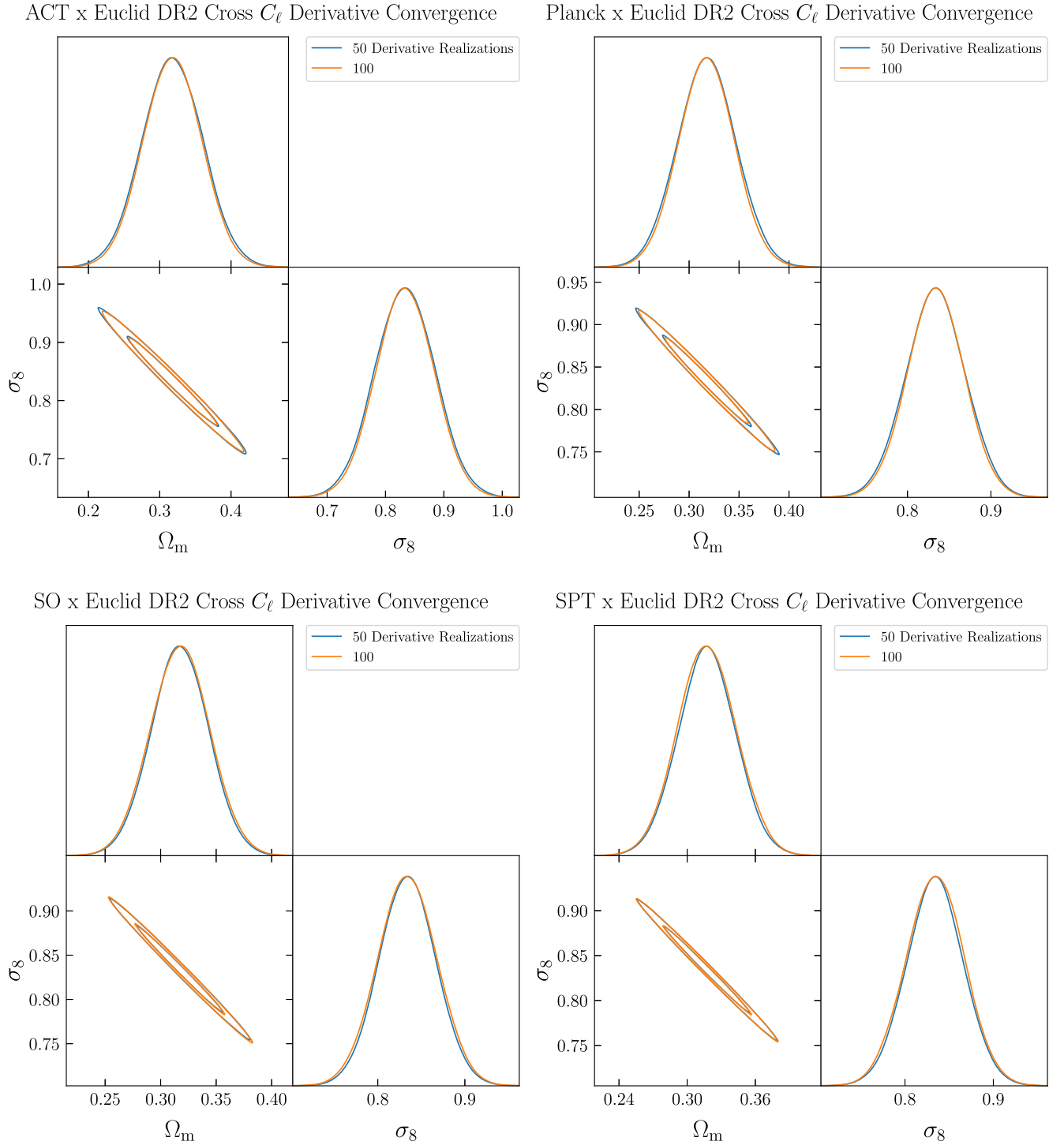


FIG. 11. Contour convergence as fewer derivatives are used for cross- C_ℓ 's applied to $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$. Shown are contours when 50% (blue) and 100% (orange) of the total available derivatives are used.

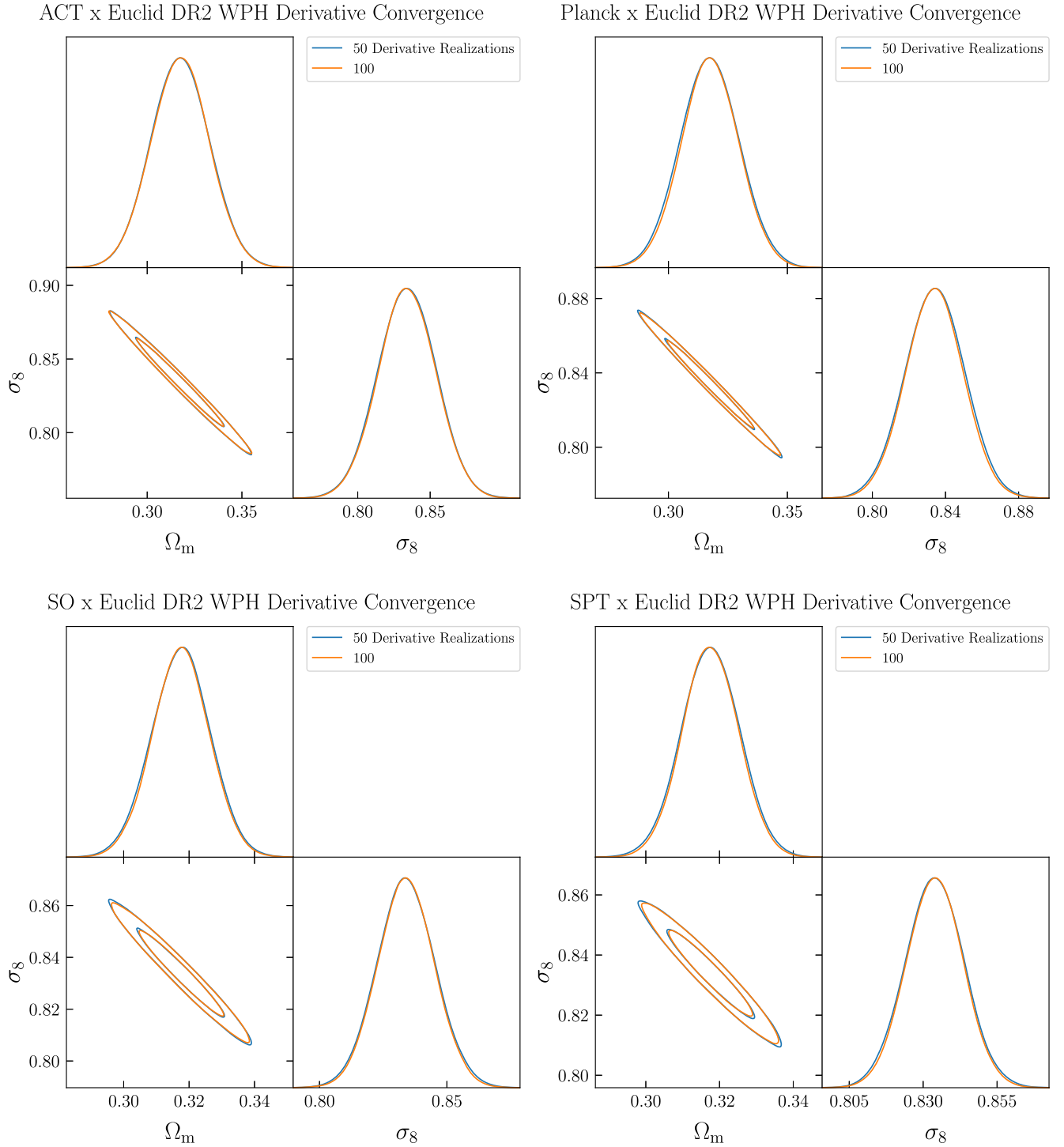


FIG. 12. Contour convergence as fewer derivatives are used for WPH applied to $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$. Shown are contours when 50% (blue) and 100% (orange) of the total available derivatives are used.

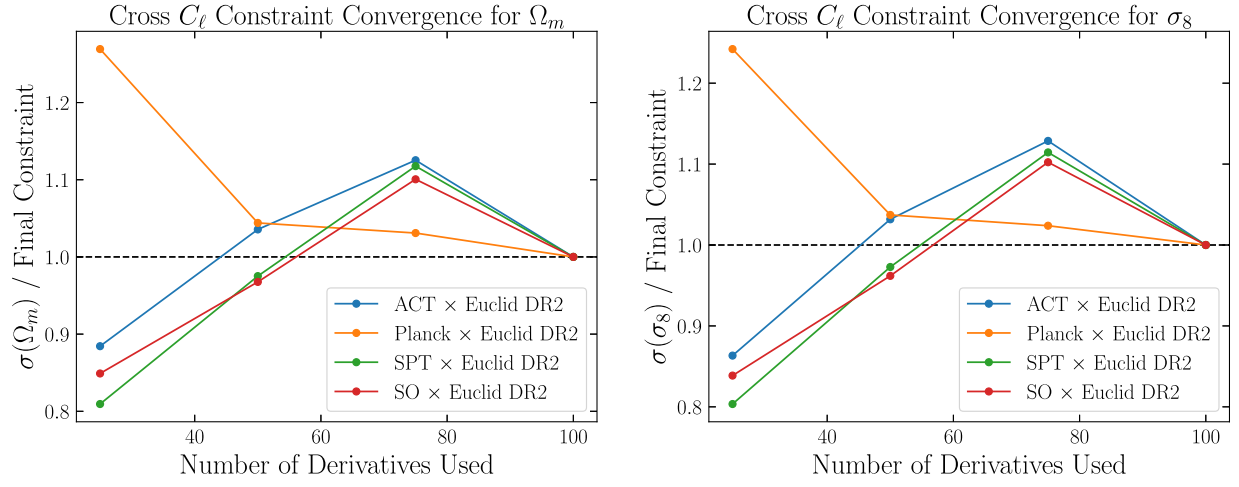


FIG. 13. Convergence of Ω_m (left) and σ_8 (right) constraints as a function of the number of derivative realizations used to evaluate cross- C_ℓ 's from $\kappa_{\text{CMB}} \times \kappa_{\text{WL}}$. By the time 50 realizations are used, constraints have converged to within $\sim 15\%$ of their final values. The high degeneracy of the parameters makes the two plots look nearly identical.

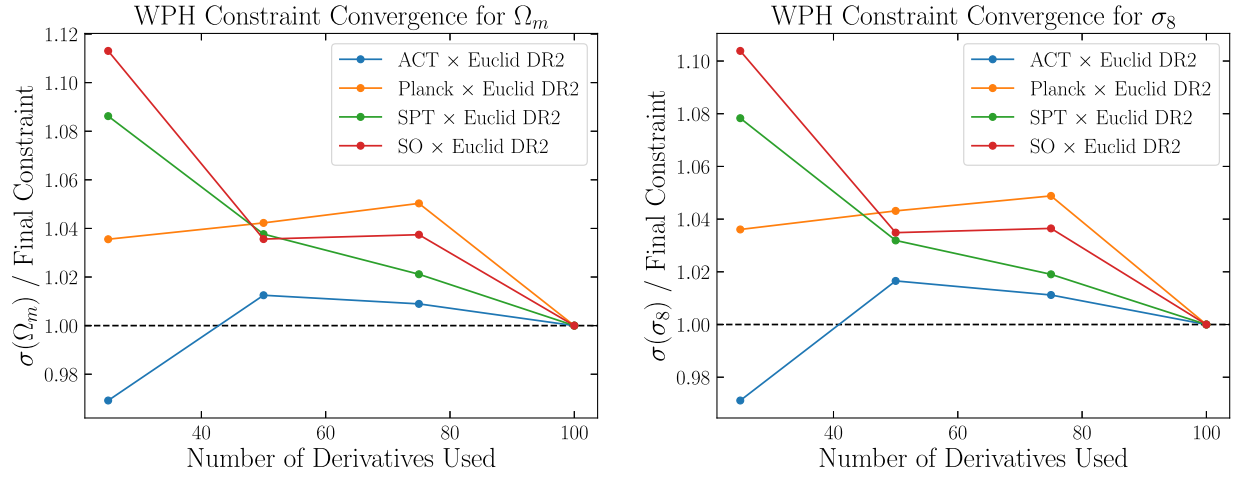


FIG. 14. Same as Fig. 13, but using WPH as the summary statistic. By the time 50 realizations are used, constraints have converged to within $\sim 5\%$ of their final values.

APPENDIX C: DERIVATIVE CONVERGENCE

This section focuses only on the convergence of derivatives of our summary statistics with respect to Ω_m . This is done for the sake of brevity; σ_8 derivatives converge to similar levels. We have sets of maps for three different values of Ω_m : the fiducial value and the fiducial value $\pm \Delta\Omega_m$ with $\Delta\Omega_m = 0.01$. In this work, for any summary statistic S , we calculate our derivatives as,

$$\frac{dS}{d\Omega_m} = \frac{\langle S(\Omega_{m,\text{fid}} + \Delta\Omega_m) - S(\Omega_{m,\text{fid}} - \Delta\Omega_m) \rangle}{2\Delta\Omega_m}, \quad (\text{C1})$$

where the average is done over the 100 realizations of these maps. Each pair of maps with $\pm \Delta\Omega_m$ shares the same initial seeding, leading to low shot-noise levels due to cancellation. We will denote this derivative estimate as S' .

Alternatively, we could have defined the derivative as,

$$\frac{dS}{d\Omega_m} = \frac{\langle S(\Omega_{m,\text{fid}}) \rangle - \langle S(\Omega_{m,\text{fid}} - \Delta\Omega_m) \rangle}{\Delta\Omega_m}, \quad (\text{C2})$$

where the averages are now over each individual set of realizations, since these maps do not share the initial seed. We will denote this derivative estimate as $\tilde{S}'$.

If our value of $\Delta\Omega_m$ is small enough, these two derivative estimates should agree to within statistical uncertainty. For statistical uncertainty, we approximate S' as being noiseless due to the shot noise cancellations coming from identical initial seeding. For $\tilde{S}'$, the majority of the shot noise will come from $\langle S(\Omega_{m,\text{fid}} - \Delta\Omega_m) \rangle$ since there are only 100 realizations for this average, compared to the 1000 fiducial realizations. We therefore model all noise as arising from

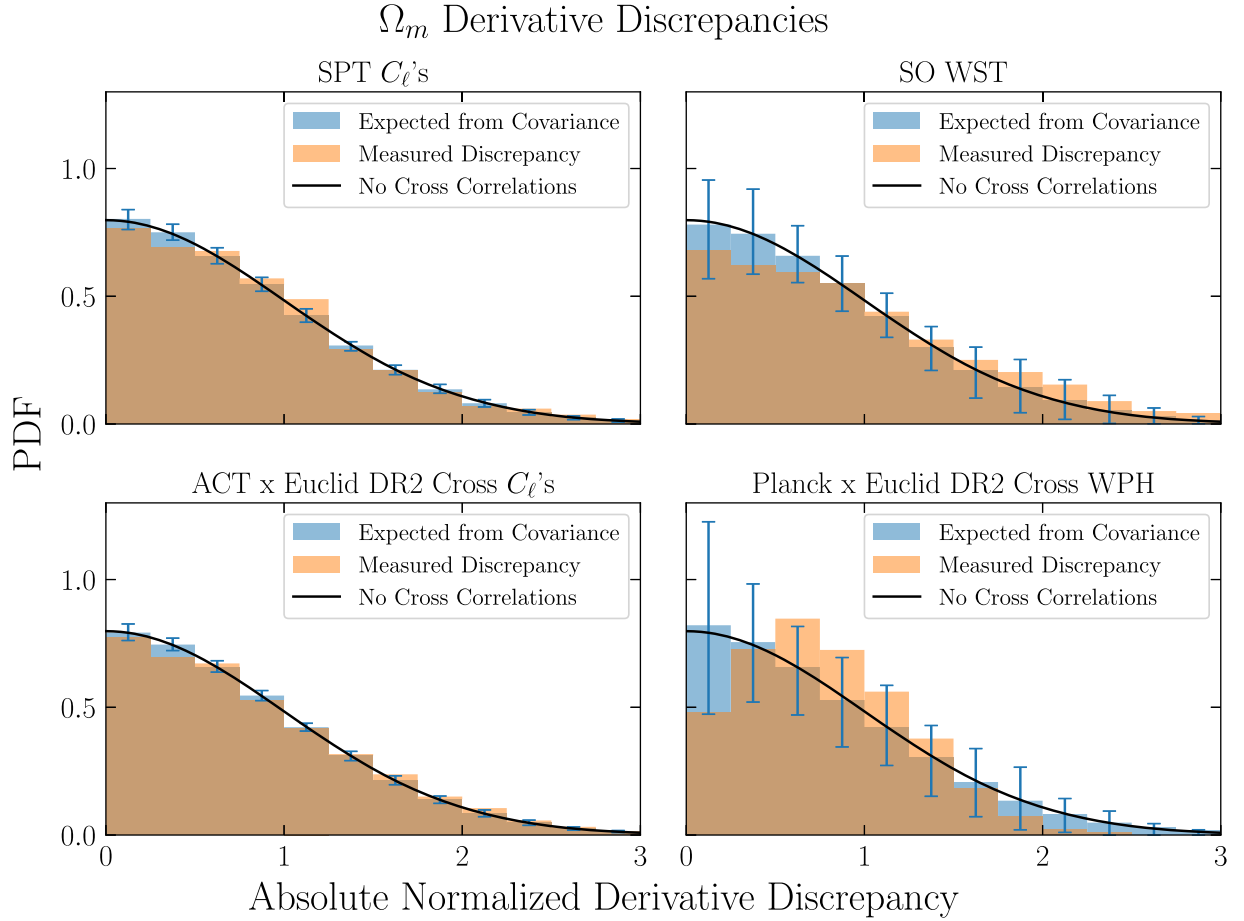


FIG. 15. Distributions of normalized derivative discrepancies with 68% error bars on histograms sampled from the fiducial covariance. One survey is picked at random for each summary statistic for brevity, but results for other surveys are comparable. Additionally, only Ω_m derivatives are shown for brevity, but σ_8 derivatives converge to comparable levels.

this term and assume that the maps are sampled from the fiducial covariance. Therefore, $\Delta\Omega_m(S' - \tilde{S}')$ should vanish to within the noise of $\langle S(\Omega_{m, \text{fid}} - \Delta\Omega_m) \rangle$. To test this, we plot histograms of,

$$\Delta\Omega_m(S' - \tilde{S}')/(\sigma(S)/\sqrt{100}), \quad (\text{C3})$$

where $\sigma(S)$ is the elementwise standard deviation of the summary statistic estimated from our fiducial realizations. The factor of $\sqrt{100}$ accounts for the majority of the noise coming from $\langle S(\Omega_{m, \text{fid}} - \Delta\Omega_m) \rangle$, which is an average of 100 realizations. If the summary statistics have no cross correlations, Eq. (C3) values should follow a Gaussian distribution. To account for cross correlations, we sample 100 vectors from the covariance matrix of the fiducial realizations with mean zero and average them elementwise. We will call this averaged sample vector D and plot out histograms of,

$$D/(\sigma(S)/\sqrt{100}). \quad (\text{C4})$$

If our derivative estimates are consistent with each other, the distributions from Eqs. (C3) and (C4) should match. We show examples of the absolute values of these histograms with 68% error bars on the sampled histogram in Fig. 15. We find no significant differences between histograms for any summary statistic across surveys, so we conclude that our derivative estimates have sufficiently converged.

APPENDIX D: LEARNED BINNING DETAILS

In order to properly construct a binning matrix, the function $f(x): [1, M] \rightarrow [1, N]$ defined in Sec. IV must satisfy certain conditions. Since the first entry of the summary statistic is mapped to the first bin, we must have $f(1) = 1$. Similarly, requiring that the last entry of the summary statistic is mapped to the last bin gives the requirement $f(M) = N$. Bins should consist of entries in the summary statistic that are adjacent to one another (no gaps in the bins), which requires that $f(x)$ be a non-decreasing function. Finally, we want to ensure there are no empty bins. There are numerous constraints one could

impose on $f(x)$ to guarantee this property, and in this work we choose to require $f'(x) \leq 1$.

To construct a binning matrix, we would generally have the m^{th} column of the binning matrix be nonzero (and equal to 1) only at the integer index closest to $f(m)$. In this work, we slightly generalize binning to allow for one summary statistic entry to be mapped into two adjacent bins. For example, if $f(2) = 3.4$, the second summary statistic entry would be divided 60% into bin 3 and 40% into bin 4. This splitting ensures that the determinant of the Fisher matrix is a continuous function of pivot positions.

As mentioned in Sec. IV, our function f is a linear interpolation between pivots, so that we perform optimization in a low-dimensional space. Once we have picked a number of pivots, our loss function is the negative logarithm of the determinant of the Fisher matrix. We include a prior term in our loss function to ensure that the pivots satisfy the conditions explained earlier in this section. We optimize this loss function using the differential evolution method in SciPy [51] with a population size

of 20. The population is initially randomly seeded throughout the pivot space.

APPENDIX E: GAUSSIANTY OF THE LIKELIHOOD

The Fisher forecast methods we use rely on our summary statistics, which follow a Gaussian distribution. With enough binning, the central limit theorem will ensure that our statistics are indeed Gaussian. We test whether this is the case by producing a χ^2 distribution of the statistics as in Valogiannis *et al.* [52]. Specifically, if a summary statistic on any given realization is given by $\mathbf{O}_i$ with an average value of $\bar{\mathbf{O}}$ and covariance matrix C , the χ^2 values can be given by,

$$\chi^2 = (\mathbf{O}_i - \bar{\mathbf{O}})^T C^{-1} (\mathbf{O}_i - \bar{\mathbf{O}}). \quad (\text{E1})$$

We perform this test with our binned data. Specifically, we use 500 fiducial realizations to learn the binning,

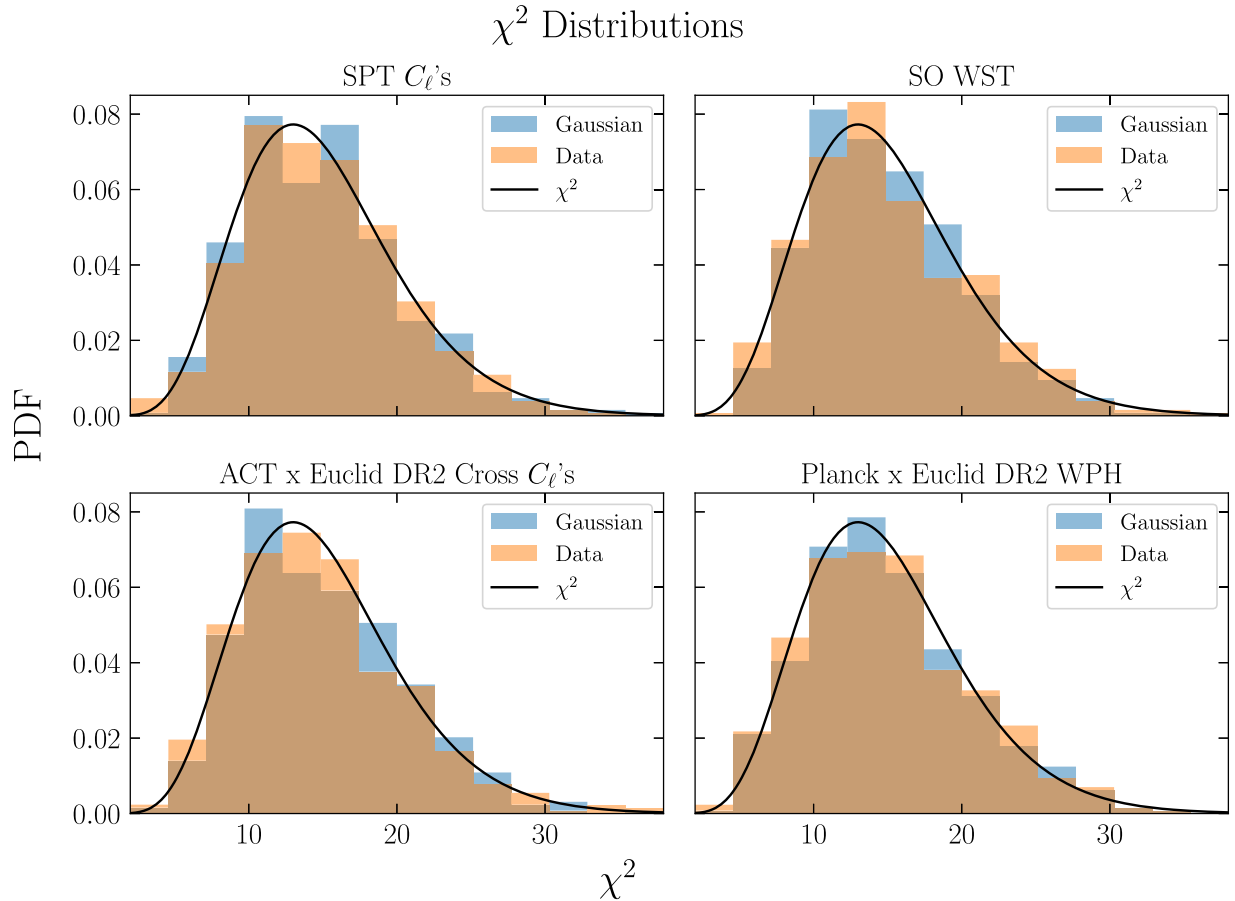


FIG. 16. Probability density functions of the χ^2 distribution for all summary statistics used. One survey is picked at random for each summary statistic for brevity, but results for other surveys are comparable. 500 fiducial realizations (orange) are used for the distributions after being binned with a binning learned by the other 500 fiducial simulations. Alongside the data distribution, we show the theoretical χ^2 distribution with 15 degrees of freedom (black) and samples drawn from a Gaussian distribution with the same mean and covariance as the data (blue). We find no evidence of significant departures from Gaussianity for any of our statistics.

then apply that binning to the remaining 500 realizations to obtain χ^2 values. Since we use 15 bins, if our statistics are Gaussian distributed, the calculated χ^2 values should follow the theoretical χ^2 distribution with 15 degrees of freedom. We verify this in Fig. 16 where

we find no significant differences between the χ^2 distribution of our data, the theoretical χ^2 distribution, and a Gaussian distribution from samples randomly generated with the same mean and covariance as our data.

-
- [1] A. Einstein, Lens-like action of a star by the deviation of light in the gravitational field, *Science* **84**, 506 (1936).
 - [2] C. Chang, V. Vikram, B. Jain, D. Bacon, A. Amara, M. R. Becker, G. Bernstein, C. Bonnett, S. Bridle, D. Brout, M. Busha, J. Frieman, E. Gaztanaga, W. Hartley, M. Jarvis, T. Kacprzak *et al.*, Wide-field lensing mass maps from dark energy survey science verification data, *Phys. Rev. Lett.* **115**, 051301 (2015).
 - [3] H. Bretonnière, U. Kuchner, M. Huertas-Company, E. Merlin, M. Castellano, D. Tuccillo, F. Buitrago, C. J. Conselice, A. Boucaud, B. Häußler, M. Kümmel, W. G. Hartley, A. Alvarez Ayllon, E. Bertin, F. Ferrari, L. Ferreira *et al.*, Euclid preparation: XXVI. The *Euclid* morphology challenge: Towards structural parameters for billions of galaxies, *Astron. Astrophys.* **671**, A102 (2023).
 - [4] DES Collaboration, The Dark Energy Survey: More than dark energy—An overview, *Mon. Not. R. Astron. Soc.* **460**, 1270 (2016).
 - [5] K. Kuijken, C. Heymans, H. Hildebrandt, R. Nakajima, T. Erben, J. T. A. de Jong, M. Viola, A. Choi, H. Hoekstra, L. Miller, E. van Uitert, A. Amon, C. Blake, M. Brouwer, A. Buddendiek, I. F. Conti *et al.*, Gravitational lensing analysis of the kilo-degree survey, *Mon. Not. R. Astron. Soc.* **454**, 3500 (2015).
 - [6] H. Aihara, N. Arimoto, R. Armstrong, S. Arnouts, N. A. Bahcall, S. Bickerton, J. Bosch, K. Bundy, P. L. Capak, J. H. H. Chan, M. Chiba, J. Coupon, E. Egami, M. Enoki, F. Finet, H. Fujimori *et al.*, The Hyper Suprime-CAM SSP survey: Overview and survey design, *Publ. Astron. Soc. Jpn.* **70**, S4 (2017).
 - [7] P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, S. Bailey, D. R. Ballantyne, J. R. Bankert, W. A. Barkhouse, J. D. Barr, L. F. Barrientos *et al.* (LSST Science and LSST Project Collaborations), LSST science book, version 2.0, [arXiv:0912.0201](https://arxiv.org/abs/0912.0201).
 - [8] M. S. Madhavacheril, F. J. Qu, B. D. Sherwin, N. MacCrann, Y. Li, I. Abril-Cabezas, P. A. R. Ade, S. Aiola, T. Alford, M. Amiri, S. Amodeo, R. An, Z. Atkins, J. E. Austermann, N. Battaglia, E. S. Battistelli *et al.*, The Atacama Cosmology Telescope: DR6 gravitational lensing map and cosmological parameters, *Astrophys. J.* **962**, 113 (2024).
 - [9] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock *et al.*, Planck 2018 results: VI. Cosmological parameters, *Astron. Astrophys.* **641**, A6 (2020).
 - [10] S. Aiola, E. Calabrese, L. Maurin, S. Naess, B. L. Schmitt, M. H. Abitbol, G. E. Addison, P. A. R. Ade, D. Alonso, M. Amiri, S. Amodeo, E. Angile, J. E. Austermann, T. Baidon, N. Battaglia, J. A. Beall *et al.*, The Atacama Cosmology Telescope: DR4 maps and cosmological parameters, *J. Cosmol. Astropart. Phys.* **12** (2020) 047.
 - [11] G. Simard, Y. Omori, K. Aylor, E. J. Baxter, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H.-M. Cho, R. Chown, T. M. Crawford, A. T. Crites, T. d. Haan, M. A. Dobbs, W. B. Everett, E. M. George *et al.*, Constraints on cosmological parameters from the angular power spectrum of a combined 2500 deg² SPT-SZ and Planck gravitational lensing map, *Astrophys. J.* **860**, 137 (2018).
 - [12] P. Ade, J. Aguirre, Z. Ahmed, S. Aiola, A. Ali, D. Alonso, M. A. Alvarez, K. Arnold, P. Ashton, J. Austermann, H. Awan, C. Baccigalupi, T. Baidon, D. Barron, N. Battaglia, R. Battye *et al.*, The Simons Observatory: Science goals and forecasts, *J. Cosmol. Astropart. Phys.* **02** (2019) 056.
 - [13] S. Cheng, Y.-S. Ting, B. Ménard, and J. Bruna, A new approach to observational cosmology using the scattering transform, *Mon. Not. R. Astron. Soc.* **499**, 5902 (2020).
 - [14] G. Valogiannis and C. Dvorkin, Going beyond the galaxy power spectrum: An analysis of boss data with wavelet scattering transforms, *Phys. Rev. D* **106**, 103509 (2022).
 - [15] G. Valogiannis and C. Dvorkin, Towards an optimal estimation of cosmological parameters with the wavelet scattering transform, *Phys. Rev. D* **105**, 103534 (2022).
 - [16] G. Valogiannis, S. Yuan, and C. Dvorkin, Precise cosmological constraints from boss galaxy clustering with a simulation-based emulator of the wavelet scattering transform, *Phys. Rev. D* **109**, 103503 (2024).
 - [17] M. Gatti, N. Jeffrey, L. Whiteway, J. Williamson, B. Jain, V. Ajani, D. Anbajagane, G. Giannini, C. Zhou, A. Porredon, J. Prat, M. Yamamoto, J. Blazek, T. Kacprzak, S. Samuroff, A. Alarcon *et al.*, Dark Energy Survey Year 3 results: Simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps I: Validation on simulations, *Phys. Rev. D* **109**, 063534 (2024).
 - [18] S. Cheng, G. A. Marques, D. Grandón, L. Thiele, M. Shirasaki, B. Ménard, and J. Liu, Cosmological constraints from weak lensing scattering transform using HSC Y1 data, *J. Cosmol. Astropart. Phys.* **01** (2025) 006.

- [19] H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, The power spectrum and bispectrum of SDSS DR11 BOSS galaxies—I. Bias and gravity, *Mon. Not. R. Astron. Soc.* **451**, 539 (2015).
- [20] O. H. E. Philcox, J. Hou, and Z. Slepian, A first detection of the connected 4-point correlation function of galaxies using the BOSS CMASS sample, [arXiv:2108.01670](#).
- [21] M. M. Ivanov, O. H. E. Philcox, T. Nishimichi, M. Simonović, M. Takada, and M. Zaldarriaga, Precision analysis of the redshift-space galaxy bispectrum, *Phys. Rev. D* **105**, 063512 (2022).
- [22] A. Chudaykin, M. M. Ivanov, and O. H. E. Philcox, Reanalyzing DESI DR1: 1. Λ CDM constraints from the power spectrum and bispectrum, [arXiv:2507.13433](#).
- [23] S. Sugiyama, R. C. H. Gomes, and B. Jain, Cosmology from a joint analysis of second and third order shear statistics with Subaru Hyper Suprime-Cam Year 3 data, *Phys. Rev. D* **112**, 123545 (2025).
- [24] E. Paillas *et al.*, Cosmological constraints from density-split clustering in the BOSS CMASS galaxy sample, *Mon. Not. R. Astron. Soc.* **531**, 898 (2024).
- [25] W. Liu, E. Paillas, C. Cuesta-Lazaro, G. Valogiannis, and W. Fang, Cosmological constraints from the Minkowski functionals of the BOSS CMASS galaxy sample, *J. Cosmol. Astropart. Phys.* **05**, 064 (2024).
- [26] J. Prat, M. Gatti, C. Doux, P. Pranav, C. Chang, N. Jeffrey, L. Whiteway, D. Anbajagane, S. Sugiyama, A. Thomsen, A. Alarcon, A. Amon, K. Bechtol, G. M. Bernstein, A. Campos, R. Chen *et al.*, Dark Energy Survey Year 3 results: w CDM cosmology from simulation-based inference with persistent homology on the sphere, *Mon. Not. R. Astron. Soc.* **545**, staf2152 (2026).
- [27] J. Armijo, G. A. Marques, C. P. Novaes, L. Thiele, J. A. Cowell, D. Grandón, M. Shirasaki, and J. Liu, Cosmological constraints using Minkowski functionals from the first year data of the Hyper Suprime-CAM, *Mon. Not. R. Astron. Soc.* **537**, 3553 (2025).
- [28] J. Fluri, T. Kacprzak, A. Lucchi, A. Refregier, A. Amara, T. Hofmann, and A. Schneider, Cosmological constraints with deep learning from KIDS-450 weak lensing maps, *Phys. Rev. D* **100**, 063514 (2019).
- [29] N. Jeffrey, L. Whiteway, M. Gatti, J. Williamson, J. Alsing, A. Porredon, J. Prat, C. Doux, B. Jain, C. Chang, T. Y. Cheng, T. Kacprzak, P. Lemos, A. Alarcon, A. Amon, K. Bechtol *et al.*, Dark Energy Survey Year 3 results: Likelihood-free, simulation-based w CDM inference with neural compression of weak-lensing map statistics, *Mon. Not. R. Astron. Soc.* **536**, 1303 (2025).
- [30] D. Anbajagane, C. Chang, H. Lee, and M. Gatti, Primordial non-Gaussianities with weak lensing: Information on non-linear scales in the ULAGAM full-sky simulations, *J. Cosmol. Astropart. Phys.* **03**, 062 (2024).
- [31] D. Potter and J. Stadel, PKDGRAV3: Beyond trillion particle cosmological simulations for the next era of galaxy surveys, *Comput. Astrophys. Cosmol.* **4**, 2 (2017).
- [32] F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys, A. Brochard, C. Uhlemann, C.-T. Chiang, S. He, A. Pisani, A. Obuljen *et al.*, The Quijote simulations, *Astrophys. J. Suppl. Ser.* **250**, 2 (2020).
- [33] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, *Astrophys. J.* **622**, 759 (2005).
- [34] Y. Omori, Agora: Multi-component simulation for cross-survey science, *Mon. Not. R. Astron. Soc.* **530**, 5030 (2024).
- [35] K. Prabhu, S. Raghunathan, M. Millea, G. Lynch, P. A. R. Ade, E. Anderes, A. J. Anderson, B. Ansarinejad, M. Archipley, L. Balkenhol, K. Benabed, A. N. Bender, B. A. Benson, F. Bianchini, L. E. Bleem, F. R. Bouchet *et al.*, Testing the Λ CDM cosmological model with forthcoming measurements of the cosmic microwave background with SPT-3G, *Astrophys. J.* **973**, 4 (2024).
- [36] W. Hu and T. Okamoto, Mass reconstruction with cosmic microwave background polarization, *Astrophys. J.* **574**, 566 (2002).
- [37] J. Liu, J. C. Hill, B. D. Sherwin, A. Petri, V. Böhm, and Z. Haiman, CMB lensing beyond the power spectrum: Cosmological constraints from the one-point probability distribution function and peak counts, *Phys. Rev. D* **94**, 103501 (2016).
- [38] S. Mallat, Group invariant scattering, *Commun. Pure Appl. Math.* **65**, 1331 (2012).
- [39] A. B. Romeo, C. Horellou, and J. Bergh, N-body simulations with two-orders-of-magnitude higher performance using wavelets, *Mon. Not. R. Astron. Soc.* **342**, 337 (2003).
- [40] S. Cheng and B. Ménard, How to quantify fields or textures? A guide to the scattering transform, [arXiv:2112.01288](#).
- [41] B. Regalado-Saint Blancard, E. Allys, F. Boulanger, F. Levrier, and N. Jeffrey, A new approach for the statistical denoising of Planck interstellar dust polarization data, *Astron. Astrophys.* **649**, L18 (2021).
- [42] J. Hartlap, P. Simon, and P. Schneider, Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix, *Astron. Astrophys.* **464**, 399 (2006).
- [43] W. R. Coulton and B. D. Wandelt, How to estimate Fisher information matrices from simulations, [arXiv:2305.08994](#).
- [44] A. F. Heavens, R. Jimenez, and O. Lahav, Massive lossless data compression and multiple parameter estimation from galaxy spectra, *Mon. Not. R. Astron. Soc.* **317**, 965 (2000).
- [45] M. Park, M. Gatti, and B. Jain, Dimensionality reduction techniques for statistical inference in cosmology, *Phys. Rev. D* **111**, 063523 (2025).
- [46] A. O. Szewciw, G. D. Beltz-Mohrmann, A. A. Berlind, and M. Sinha, Toward accurate modeling of galaxy clustering on small scales: Constraining the galaxy-halo connection with optimal statistics, *Astrophys. J.* **926**, 15 (2022).
- [47] J. D. McEwen, C. G. R. Wallis, and A. N. Mavor-Parker, Scattering networks on the sphere for scalable and rotationally equivariant spherical CNNs, [arXiv:2102.02828](#).
- [48] J. D. McEwen, P. Vielva, Y. Wiaux, R. B. Barreiro, L. Cayon, M. P. Hobson, A. N. Lasenby, E. Martinez-Gonzalez, and J. L. Sanz, Cosmological applications of a wavelet analysis on the sphere, *J. Fourier Anal. Appl.* **13**, 495 (2007).

- [49] K.M. Surrao and J.C. Hill, Constraining cosmological parameters with needlet internal linear combination maps II: Likelihood-free inference on NILC power spectra, *Phys. Rev. D* **110**, 063510 (2024).
- [50] <http://iaifi.org/>.
- [51] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).
- [52] G. Valogiannis, F. Villaescusa-Navarro, and M. Baldi, Towards unveiling the large-scale nature of gravity with the wavelet scattering transform, *J. Cosmol. Astropart. Phys.* **11** (2024) 061.