BNL-225600-2024-JAAM

Dark Energy Survey Year 3 results: Simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps. Validation on simulations

M. Gatti, E. Sheldon

To be published in "Physical Review D"

March 2024

Physics Department

Brookhaven National Laboratory

U.S. Department of Energy

USDOE Office of Science (SC), High Energy Physics (HEP)

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No.DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DES-2023-0634
FERMILAB-PUB-23-634-PPD

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

M. Gatti,[1,*] N. Jeffrey,[2] L. Whiteway,[2] J. Williamson,[2] B. Jain,[1] V. Ajani,[3] D. Anbajagane,[4,5] G. Giannini,[4] C. Zhou,[6] A. Porredon,[7] J. Prat,[4,5] M. Yamamoto,[8] J. Blazek,[9] T. Kacprzak,[3] S. Samuroff,[9] A. Alarcon,[10] A. Amon,[11,12] K. Bechtol,[13] M. Becker,[14] G. Bernstein,[1] A. Campos,[13] C. Chang,[4,5] R. Chen,[8] A. Choi,[15] C. Davis,[16] J. Derose,[17] H. T. Diehl,[18] S. Dodelson,[19,20] C. Doux,[21] K. Eckert,[1] J. Elvin-Poole,[22] S. Everett,[23] A. Ferte,[24] D. Gruen,[25] R. Gruendl,[26,27] I. Harrison,[28] W. G. Hartley,[29] K. Herner,[18] E. M. Huff,[23] M. Jarvis,[1] N. Kuropatkin,[18] P. F. Leget,[16] N. MacCrann,[30] J. McCullough,[16] J. Myles,[31] A. Navarro-Alsina,[32] S. Pandey,[1] M. Raveri,[33] R. P. Rollins,[34] A. Roodman,[16,24] C. Sanchez,[1] L. F. Secco,[4] I. Sevilla-Noarbe,[35] E. Sheldon,[36] T. Shin,[37] M. Troxel,[8] I. Tutusaus,[38] T. N. Varga,[39,40,41] B. Yanny,[18] B. Yin,[19] Y. Zhang,[42] J. Zuntz,[43] M. Aguena,[44] O. Alves,[45] J. Annis,[18] D. Brooks,[2] J. Carretero,[46] F. J. Castander,[47] R. Cawthon,[13] M. Costanzi,[48] L. N. da Costa,[44,49] M. E. S. Pereira,[50] A. E. Evrard,[45,51] B. Flaugher,[18] P. Fosalba,[47] J. Frieman,[18] J. García-Bellido,[52] D. W. Gerdes,[45,51] D. Gruen,[24,16] R. A. Gruendl,[53,54] J. Gschwend,[44,49] G. Gutierrez,[18] D. L. Hollowood,[55] K. Honscheid,[56] D. J. James,[57] K. Kuehn,[58,59] O. Lahav,[2] S. Lee,[23] J. L. Marshall,[60] J. Mena-Fernández,[61] F. Menanteau,[26,27] R. Miquel,[62,63] R. L. C. Ogando,[64] M. E. S. Pereira,[65] A. Pieres,[44,64] A. A. Plazas Malagón,[16,24] E. Sanchez,[35] M. Smith,[66] E. Suchyta,[67] M. E. C. Swanson,[26] G. Tarle,[45] N. Weaverdyck,[45,68] J. Weller,[40,69] and P. Wiseman[66]

(DES Collaboration)

[1] *Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[2] *Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK*
[3] *Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[4] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[5] *Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
[6] *Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[7] *Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute, German Centre for Cosmological Lensing, 44780*
[8] *Department of Physics, Duke University Durham, NC 27708, USA*
[9] *Department of Physics, Northeastern University, Boston, MA 02115, USA*
[10] *Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*
[11] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[12] *Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[13] *Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390*
[14] *Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*
[15] *NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA*
[16] *Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, CA 94305, USA*
[17] *Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[18] *Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA*
[19] *Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA*
[20] *NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[21] *Université Grenoble Alpes, CNRS, LPSC-IN2P3, 38000 Grenoble, France*
[22] *Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*
[23] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[24] *SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[25] *University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 Munich, Germany*
[26] *Center for Astrophysical Surveys, National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*
[27] *Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[28] *School of Physics and Astronomy, Cardiff University, CF24 3AA, UK*
[29] *Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland*
[30] *Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*
[31] *Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[32] *Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil*
[33] *Department of Physics, University of Genova and INFN, Via Dodecaneso 33, 16146, Genova, Italy*

arXiv:2310.17557v2 [astro-ph.CO] 4 Nov 2023

[34] Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[35] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[36] Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA
[37] Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA
[38] Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
[39] Excellence Cluster Origins, Boltzmannstr. 2, 85748 Garching, Germany
[40] Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany
[41] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany
[42] Cerro Tololo Inter-American Observatory, NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile
[43] Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK
[44] Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[45] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[46] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain
[47] Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain
[48] INAF-Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34131 Trieste, Italy
[49] Observatório Nacional, Rua General José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[50] Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany
[51] Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
[52] Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, Cantoblanco, Madrid 28049, Spain
[53] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA
[54] National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA
[55] Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, CA 95064, USA
[56] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
[57] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[58] Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia
[59] Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA
[60] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[61] LPSC Grenoble - 53, Avenue des Martyrs 38026 Grenoble, France
[62] Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
[63] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain
[64] Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[65] Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany
[66] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
[67] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[68] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[69] Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 München, Germany

Beyond-two-point statistics contain additional information on cosmological as well as astrophysical and observational (systematics) parameters. In this methodology paper we provide an end-to-end simulation-based analysis of a set of Gaussian and non-Gaussian weak lensing statistics using detailed mock catalogues of the Dark Energy Survey. We implement: 1) second and third moments; 2) wavelet phase harmonics (WPH); 3) the scattering transform (ST). Our analysis is fully based on simulations, it spans a space of seven νwCDM cosmological parameters, and it forward models the most relevant sources of systematics of the data (masks, noise variations, clustering of the sources, intrinsic alignments, and shear and redshift calibration). We implement a neural network compression of the summary statistics, and we estimate the parameter posteriors using a likelihood-free-inference approach. We validate the pipeline extensively, and we find that WPH exhibits the strongest performance when combined with second moments, followed by ST, and then by third moments. The combination of all the different statistics further enhances constraints with respect to second moments, up to 25 per cent, 15 per cent, and 90 per cent for S_8, $\Omega_{\rm m}$, and the Figure-Of-Merit $\mathrm{FoM}_{S_8,\Omega_{\rm m}}$, respectively. We further find that non-Gaussian statistics improve constraints on w and on the amplitude of intrinsic alignment with respect to second moments constraints. The methodological advances presented here are suitable for application to Stage IV surveys from

Euclid, Rubin-LSST, and Roman with additional validation on mock catalogues for each survey. In a companion paper we present an application to DES Year 3 data.

I. INTRODUCTION

Weak gravitational lensing is a powerful tool for studying the large-scale structure (LSS) of the mass distribution in the Universe. Photons emitted by distant galaxies are deflected when passing through regions of spacetime affected by the mass distribution between the sources and the observer [29]. By measuring the shapes of numerous galaxies, statistical methods enable us to deduce the projected spatial distribution of the mass responsible for these weak deflections and thereby create weak lensing *mass maps* (Chang et al. 19, 20, Liu et al. 68, Oguri et al. 79, Van Waerbeke et al. 106, Vikram et al. 110; Jeffrey & Gatti et al. 57). At the time of writing, ongoing and upcoming surveys, including the Dark Energy Survey (DES, Collaboration 23), the Kilo-Degree Survey (KIDS, Kuijken et al. 65), the Hyper Suprime-Cam (HSC, Aihara et al. 3), the Vera C. Rubin Observatory's Legacy Survey [69], and the Euclid mission [66], are measuring (or being readied to measure) galaxy shapes on a massive scale, encompassing thousands of square degrees across the sky. Notably, the DES project recently measured the shapes of more than 100 million galaxies in an area of approximately 5000 square degrees in the southern hemisphere [43], which led to the production of the most extensive weak lensing mass map from a galaxy survey to date [57]. In parallel, measurements of the lensing of the cosmic microwave background (CMB) have led to maps of the mass distribution projected all the way to the redshift of the last scattering surface (e.g. Madhavacheril et al. 72).

If a mean-zero random field is Gaussian, then a two-point statistic captures all its statistical information. Two-point statistics of the shear field can be measured in harmonic, configuration, or other spaces: e.g. power spectra (harmonic space), shear two-point correlation function (configuration space), or COSEBI (Complete Orthogonal Sets of E/B-Integrals) have to date been measured and used for cosmological parameter estimation (e.g. Asgari et al. 10; Amon et al. 7; Secco & Samuroff et al. 96; Dalal et al. 26, Doux et al. 28, Li et al. 67). However, a significant amount of the information contained in weak lensing mass maps lies in their non-Gaussian features, and these features are not fully captured by two-point statistics. Many recent studies, using a wide range of tools and statistics, have tried to extract the non-Gaussian information; examples include higher-order moments [19, 39, 41, 83, 84, 86, 106–108], peak counts [4, 27, 49, 60, 62, 68, 77, 83, 97, 111, 112], one-point probability distributions [12, 16, 100], Minkowski functionals [46, 63, 81, 84, 109], Betti numbers [32, 82], persistent homology [52, 53], scattering transform coefficients [21, 102, 103], wavelet phase harmonic moments [5], kNN and CDFs [8, 11], map-level inference [15, 85], and machine-learning methods [34, 35, 56, 70, 89]. Many of these studies, however, are limited to being proofs of concept, restricted to idealized simulated scenarios (due to the challenges associated with applying these techniques to real-world data). Nevertheless, the field is rapidly progressing, with a number of recent applications to observational data [33, 34, 41, 52, 56, 60, 68, 70, 77, 111].

One of the major challenges in exploiting non-Gaussian statistics is the need for accurate modelling of measurements. Analytic models are available only for a small set of non-Gaussian summary statistics (e.g. moments), and often these models are reliable only at large scales. Consequently, many studies resort to using simulations to forward model the observables. This procedure introduces its own challenges. Most importantly, computational resources are a significant concern, as it is necessary to run numerous N-body simulations to cover the parameter space explored in the analysis. Additionally, it is a formidable task to incorporate all the relevant observational and systematic effects into these simulations. Finally, it is critical to estimate efficiently the parameter posteriors; this requires techniques able to recover accurately the posterior from a limited number of simulation samples (specifically, those available at the locations in parameter space of the N-body simulations).

In this study, we use a set of non-Gaussian summary statistics of weak lensing mass maps to constrain cosmology with the first three years (Y3) of data from DES. This work validates the methodology using simulations; a companion paper applying this framework to the DES Y3 data will follow. Our analysis makes use of the following Gaussian and non-Gaussian statistics: second and third-order moments, wavelet phase harmonic (WPH) moments, and the scattering transform (ST) coefficients. Moments have previously been used in analysing DES data using analytical models instead of simulations [41]; in contrast, this paper fully relies on a simulation based inference. Furthermore, WPH moments and the ST have not been applied to data before. The WPH moments are second moments of smoothed weak lensing mass maps that have undergone a non-linear transformation, allowing for the exploration of the non-Gaussian features of the field. The ST coefficients are built through a series of smoothing and modulus operations applied to the input field, followed by an average. WPH and ST have two advantages relative to traditional higher-order correlations: better constraining power and (as they do not go to higher than second order in the field) lower sensitivity to noise fluctuations [5]. WPH and ST are frequently compared to convolutional neural networks (CNNs) because their definition bears similarities to the architecture of CNNs [75]; however, their definition depends only on a handful of parameters (parameters that have clear physical interpretation), and, in contrast to CNNs, they

* marcogatti29@gmail.com

require no training.

For this work we produced a set of N-body simulations (Jeffrey et al., in prep.) that explores a seven-dimensional parameter space. The simulations incorporate key observational and astrophysical systematic effects impacting weak lensing analyses, including photometric redshift uncertainties, shear calibration errors, intrinsic alignments, and source clustering (as described in [42], this latter effect has a greater influence on non-Gaussian statistics than on Gaussian statistics). To obtain posterior estimates of the parameters, we employ an optimal data compression technique called neural compression, which significantly reduces the dimensionality of our summary statistics. Subsequently, we employ a likelihood-free inference (LFI, e.g. Jeffrey et al. 56) approach, enabling us to estimate posteriors without imposing restrictive assumptions about the likelihood or data model. This powerful approach circumvents various technical challenges associated with conventional analysis methods, such as covariance matrix estimation and sampling from high-dimensional Bayesian hierarchical models. We also examine the combination of the three non-Gaussian summary statistics considered in this work; to date, the combination of distinct non-Gaussian summary statistics has only been explored in idealized simulations [30, 111], and its application to real data remains unexplored. We test the methodology extensively with simulated data to ensure that the results from survey data are unbiased.

This paper is organised as follows. Section 2 summarizes the survey data as well as the simulations used for our model predictions and for validation. Section 3 describes the various summary statistics, their covariances, and the compressed statistics obtained from them. We describe and validate in Section 4 the LFI methodology for parameter inference and in Section 5 the choice of scale cuts. Section 6 validates the full pipeline with an end-to-end simulated cosmological analysis, and we summarise our results in Section 7.

II. DATA AND SIMULATIONS

A. DES Y3 weak lensing catalogue

We use the DES Y3 weak lensing catalogue [43]; this contains 100,204,026 galaxies, with a weighted $n_{\rm eff}$ = 5.59 galaxies arcmin^{-2}, over an effective area of 4139 deg^2. It was created using the METACALIBRATION algorithm [54, 98], which provides self-calibrated shear estimates starting from (multi-band) noisy images of the detected objects. A residual small calibration (in the form of a multiplicative shear bias) is provided; it is based on sophisticated image simulations [71] and it accounts for blending-related detection effects. An inverse variance weight is further assigned to each galaxy in the catalogue to enhance the overall signal-to-noise. The sample is divided into four tomographic bins of roughly equal number density [78] and redshift distributions are provided by the SOMPZ method [78] in combination with clustering redshift constraints [40] and corrections due to the redshift-dependent effects of blending [71].

Table I. Distribution of the parameters spanned by the Gower St mock catalogues (second column), and the prior used in the cosmological analysis (third column). The prior used in the analysis can differ from the distribution of the samples as long as these parameters have been explicitly used during the training of the Neural Density Estimators (NDEs) when learning the likelihood surface; more details are given in §IV. For the third column, we report the analysis prior only if it is different from the mocks parameters distribution.

Parameter	Mocks parameters distribution	Analysis prior
$\Omega_{\rm m}$	mixed active-learning in $\mathcal{U}(0.15, 0.52)$	$\mathcal{U}(0.15, 0.52)$
S_8	mixed active-learning in $\mathcal{U}(0.5, 1.0)$	$\mathcal{U}(0.5, 1.0)$
w	$\mathcal{N}(-1, \frac{1}{3})$ for $-1 < w < -\frac{1}{3}$ 0 else	$\mathcal{U}(-1, \frac{1}{3})$
n_s	$\mathcal{N}(0.9649, 0.0063)$	
h	$\mathcal{N}(0.7022, 0.0245)$	
$\Omega_{\rm b}h^2$	$N(0.02237, 0.00015)$	
$\log(m_\nu)$	$\mathcal{U}[\log(0.06), \log(0.14)]$	
A_{IA}	$\mathcal{U}[-3, 3]$	
η_{IA}	$\mathcal{U}[-5, 5]$	
m_1	$\mathcal{N}(-0.0063, 0.0091)$	
m_2	$\mathcal{N}(-0.0198, 0.0078)$	
m_3	$\mathcal{N}(-0.0241, 0.0076)$	
m_4	$\mathcal{N}(-0.0369, 0.0076)$	
$\bar{n}_i(z)$	$p_{\rm HyperRank}(\bar{n}_i(z)\|x_{\rm phot})$	

B. Simulations

1. Gower St simulations

We use the Gower St simulation suite (Jeffrey et al., in prep.) to build our pipeline. The suite consists of 791 gravity-only full-sky N-body simulations, produced using the PKDGRAV3 code [87]. The simulations span a seven-dimensional parameter space in νwCDM ($\Omega_{\rm m}$, σ_8, n_s, h_{100}, $\Omega_{\rm b}$, w, m_ν). The parameter space is not spanned uniformly (Fig. 2 shows the simulation distribution in the parameter space). $\Omega_{\rm m}$ and σ_8 have been sampled with a mixed active-learning strategy; in particular they were at first distributed according to the existing DES analysis constraints, and then, after an initial simple blind power spectrum analysis, new simulations were run with σ_8 and $\Omega_{\rm m}$ values (known only to the computer) in regions of parameter space with poor accuracy of the likelihood estimates (see Jeffrey et al., in prep.). The other parameters were chosen to be distributed as follows:

- $n_s \sim \mathcal{N}(0.9649, 0.0063)$; from Planck [2] but with the standard deviation boosted by a factor of 1.5.
- $h \sim \mathcal{N}(0.7022, 0.0245)$; consistent with both SH0ES [90] and Planck [2].

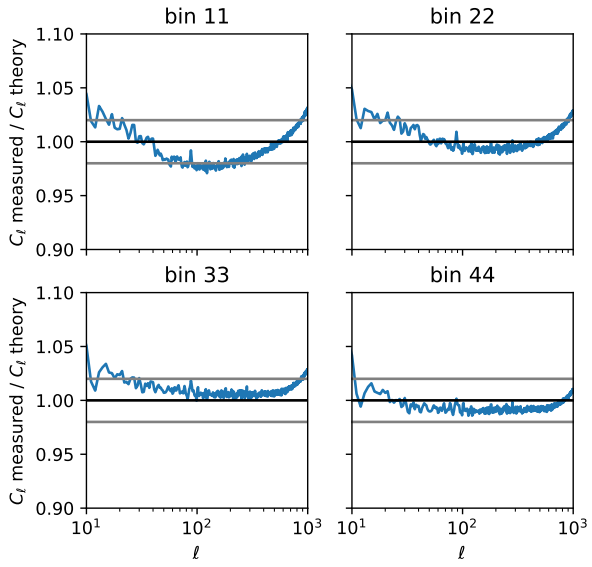

Figure 1. Ratio of the convergence power spectrum (C_ℓ) as measured in the Gower St simulations and that from theoretical predictions. The power spectrum has been measured on full-sky, noiseless convergence maps. The ratio has been averaged over all the simulations available. The two horizontal lines are provided for reference and show that the typical deviation is at the 2 percent level.

- $\Omega_b h^2 \sim N(0.02237, 0.00015)$; from Planck [2].
- $w \sim \mathcal{N}(-1, 1/3)$, but with values less than -1 or greater than $-1/3$ then discarded. For a few (64) simulations, part of the 'science verification' runs, this discarding was not done. We kept these simulations during the training of our NDEs, but we used a hard prior at $w > -1$ for the analysis.
- m_ν: fixed at 0.06 for 192 simulations and with $\log(m_\nu) \sim \mathcal{U}[\log(0.06), \log(0.14)]$ thereafter.

In the above, $\mathcal{N}(\mu, \sigma)$ denotes a normal distribution with the indicated mean and standard deviation and $\mathcal{U}[a, b]$ denotes a uniform distribution with the indicated limits. We note that the sampling strategy does not necessarily affect our posteriors; more details are given in § IV.

The simulations used up to ten replicated boxes in each direction so as to span the redshift interval from $z = 0$ to $z = 49$, although note that the bulk of our redshift distributions ($z < 1.5$) can be covered by only three replications. Each individual box contains 1080^3 particles and has a side-length of 1250 h^{-1} Mpc. For each simulation, lens planes $\delta_{\rm shell}(\hat{\mathbf{n}}, \chi)$ are provided at ~ 100 redshifts from $z = 49$ to $z = 0.0$, equally spaced in proper time. For this work, we downsample the original resolution of NSIDE = 2048 to NSIDE = 512 (with pixel size ≈ 6.9 arcmin). The lens planes are provided as HEALPix [45] maps and are obtained from the raw number particle counts.[2] The lens planes are converted into convergence planes $\kappa_{\rm shell}(\hat{\mathbf{n}}, \chi)$ under the Born approximation (e.g. Eq. 2 from Fosalba et al. 38). Lastly, shear planes $\gamma_{\rm shell}(\hat{\mathbf{n}}, \chi)$ are obtained from the convergence maps using a full-sky generalisation of the [61] algorithm [57].

We validate the Gower St simulations by comparing the power spectra measured on the full-sky convergence maps, weighted by the DES redshift distributions, against theory predictions obtained using `halofit` [99]. Note that we did not use the more recent (and more accurate) `EuclidEmu` [31] for this comparison, as `EuclidEmu` covers only a very limited portion of our parameter space. We generally do not expect an agreement better than 2 per cent, as this is the typical relative error between different non-linear power spectrum prescriptions or other modelling implementations (e.g. neutrinos). At the largest scales, on the other hand, box-size effects and/or cosmic variance in the simulations might impact the comparison. To perform the test, we build the redshift weighted convergence maps as

$$\kappa(p) = \frac{\sum_s \bar{n}(s)\kappa(p, s)}{\sum_s \bar{n}(s)}, \tag{1}$$

where p is a map pixel, s is the redshift shell, $\kappa(p, s)$ is the noiseless convergence field from the simulation, and $\bar{n}(s)$ is the DES galaxy count across the whole footprint [78]. For each of the four DES tomographic bins, we computed the ratio between the power spectrum of the simulated convergence field $\kappa(p)$ and the theory predictions from `halofit`. We show the average of the ratio over all the Gower St simulations in Fig. 1; the agreement is good, within 2 percent over the range of multipoles considered in this work (up to $\ell = 1024$; see §III).

2. *CosmoGridV1 simulations*

We use a subset of the simulations from the `CosmoGridV1` suite [59] for additional testing and to determine the scale cuts that need to be removed because of baryonic contamination. The `CosmoGridV1` simulations have been produced using the PKDGRAV3 code [87]. From the available `CosmoGridV1` simulations we chose a set of one hundred full-sky simulations at the fiducial cosmology $\sigma_8 = 0.84$, $\Omega_m = 0.26$, $w = -1$, $H_0 = 67.36$, $\Omega_b = 0.0493$, $n_s = 0.9649$. Each individual simulation has also been post-processed with a baryonification algorithm that mimics the impact of baryons at small scales. The algorithm used is the baryonic correction model [9, 93], which adjusts the particle positions in gravity-only simulations to mimic the impact of various baryonic processes on the density distribution. The cosmology has been chosen to be centred well within our

[2] $\delta_{\rm shell}(\hat{\mathbf{n}}, \chi) = n_p(\hat{\mathbf{n}}, \chi)/\langle n_p(\hat{\mathbf{n}}, \chi)\rangle - 1$, where $\langle\rangle$ indicates the spatial average and n_p is the number of particles in a given pixel p.

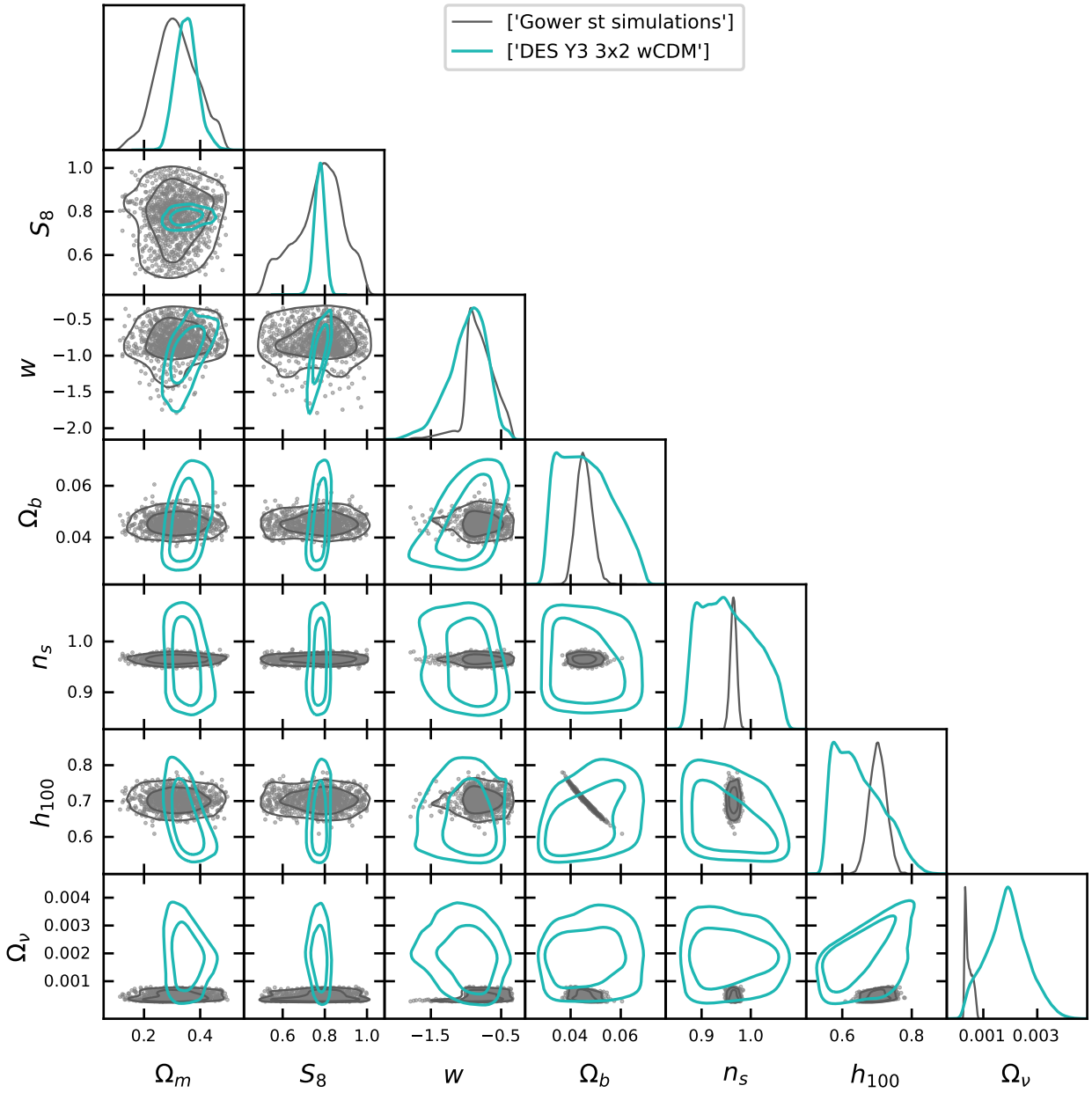

Figure 2. Distribution of Gower St simulations for the seven parameters spanned in this analysis (grey). The two-dimensional marginalised contours in these figures show the 68 per cent and 95 per cent percentile of the simulations. For comparison purposes, we also show (cyan) the posterior from the DES Y3 3x2 *vw*CDM analysis [1].

priors for σ_8, $\Omega_{\rm m}$ and w. The baryonic correction model depends on several parameters (up to seven); these impact both the shape and the amplitude of the power spectrum. The parameter that has the largest impact is M_c, the mass scale at which haloes have lost half of their gas. A value of $M_c = 10^{13.82} M_\odot$ has been adopted, following [33, 94], based on observed X-ray gas fractions. The values of the other parameters have been estimated by comparing against current X-ray observation; see [94] model `B-avrg` for a list of the values. More details are given in §V, where we evaluate the impact of baryons on our constraints.

The simulations were obtained using multiple replicated boxes in each direction so as to span the redshift interval from z = 0 to z = 3.5. Each individual box contains 832^3 particles and has a side-length of 900 h^{-1} Mpc. For each simulation, lens planes $\delta_{\rm shell}(\hat{\mathbf{n}}, \chi)$ are provided at ~ 69 redshifts from $z = 3.5$ to $z = 0.0$, equally spaced in proper time. We downsample the original resolution of NSIDE = 2048 to NSIDE = 512 (with pixel size ≈ 6.9 arcmin). Last, convergence and shear planes are obtained using the same procedure as adopted for the Gower St simulations.

3. DES Y3 maps-making procedure

We use the simulated full-sky convergence maps to generate DES Y3-like weak lensing convergence maps following the procedure outlined in [42]. The procedure is similar to others used in past DES analyses (e.g. Gatti et al. 41, Zürcher et al. 113), but improves upon them by introducing for the first time an efficient recipe to forward model source clustering effects. We further extend that procedure to incorporate extra observational systematic effects. Let p be a pixel, s a thin redshift shell, $\gamma(p,s)$ the noiseless shear from the shear simulation, and $\bar{n}(s)$ the galaxy count across the whole footprint [78]. Additionally, let m be the multiplicative shear bias that models shear measurement uncertainties [71], and let $\gamma_{\rm IA}(p,s)$ be the intrinsic alignment contribution to each pixel. Let $\delta(p,s)$ be the matter overdensity in the shear simulation, and let b_g be the galaxy-matter bias of the weak lensing sample. Each galaxy has a shear weight w_g and ellipticity e_g. We randomly rotate the DES galaxy ellipticities to erase the cosmological signal of the catalogue.

The mock shear signal in pixel p is set to

$$\gamma(p) = \frac{\sum_s \bar{n}(s)[1+b_g\delta(p,s)](1+m)[\gamma(p,s)+\gamma_{\rm IA}(p,s)]}{\sum_s \bar{n}(s)[1+b_g\delta(p,s)]} + \left(\frac{\sum_s \bar{n}(s)}{\sum_s \bar{n}(s)\left[1+b_g\delta(p,s)\right]}\right)^{1/2} F(p)\,\frac{\sum_g w_g e_g}{\sum_g w_g}. \quad (2)$$

The signal term is a weighted average over shells; here the weights have been amended to include a shear-correlated source galaxy count [42]. The term $F(p)$ in Eq. 2 is a near-unity scale factor introduced to avoid double-counting source clustering effects, adjusting the even moments of the noise of the maps, as the DES Y3 catalogue used to model the shape noise of the pixels is already affected by source clustering. We follow [42] and assume

$$F(p) = A\sqrt{1 - B\sigma_e^2(p)}, \quad (3)$$

where $\sigma_e^2(p)$ is the variance of the pixel noise and $A = [0.97, 0.985, 0.990, 0.995]$ and $B = [0.1, 0.05, 0.035, 0.035]$ are constants (one for each tomographic bin). A further validation of the noise properties of our simulations is provided in Appendix A. The intrinsic alignment term $\gamma_{\rm IA}(p,s)$ is

$$\gamma_{\rm IA}(p,s) = A_{\rm IA}\left(\frac{1+z}{1+z_0}\right)^{\eta_{\rm IA}} \frac{c_1\rho_{\rm crit}\Omega_m}{D(z)} S(p,s), \quad (4)$$

with $z_0 = 0.62$, $c_1 = 5\times10^{-14} M_\odot h^{-2}{\rm Mpc}^2$ (Bridle & King 17), $\rho_{\rm crit}$ the critical density, $D(z)$ the linear growth factor, and $S(p,s)$ the shear tidal field. We obtain $S(p,s)$ directly from the density field $\delta(p,s)$ by applying the (inverse) Kaiser-Squires algorithm. The two intrinsic alignment parameters $A_{\rm IA}$ and $\eta_{\rm IA}$ in Eq. 4 control respectively the amplitude and the redshift evolution of the intrinsic alignment signal. In writing Eq. 4 we have followed the non-linear alignment model (NLA, Bridle & King 17); however, since we are including source clustering in our simulations (the $(1+b_g)\delta(p,s)$ term in Eq. 2), the final intrinsic alignment model includes extra clustering terms beyond the original NLA implementation. These terms are similar to the clustering term included in the tidal-torque alignment (TATT) model [14]; that paper, however, estimates those contributions only for catalogue-based Gaussian statistics using tree-level perturbation theory, whereas our implementation directly uses the clustering of the simulation and generalises to all the summary statistics considered in this work. With the simulations at hand, we were not able to include a more sophisticated IA model (e.g, including all the terms of the TATT model, as was done for the fiducial DES Y3 weak lensing analysis of Amon et al. 7, Secco & Samuroff et al. 96). However, we note that the DES Y3 cosmological analyses on data (Amon et al. 7, [96], [1]) have not yielded any substantial indications favouring the adoption of a more complex model (such as TATT) over NLA; moreover, these results are consistent with a zero intrinsic alignment amplitude. For these reasons, we consider the IA model implemented here to be adequate for our analysis.

This procedure is repeated for each of the four tomographic bins of the DES Y3 source catalogue. As we can cut four independent DES Y3 footprints from each full-sky map, we produce a total of 3164 independent DES Y3 shear mock maps from the Gower St simulations. Additionally, we produced another ~9492 pseudo-

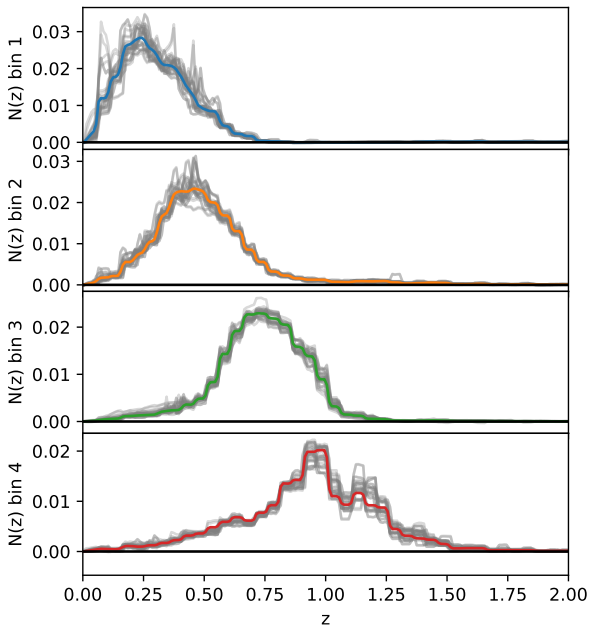

Figure 3. Redshift distributions as estimated in data for the four DES Y3 tomographic bins [78]. The solid coloured lines are the average $n(z)$ for each bin; the grey lines are a few (~ 50) samples encompassing the redshift calibration uncertainties and that we use to create the mocks.

independent DES Y3 shear mock maps by shifting the four independent DES Y3 footprints by 45, 90, and 135 degrees, so as to cover slightly different parts of the full-sky maps we generated, for a total of 12656 mocks. We used these mocks to train the neural network compression of the summary statistics. Then we repeated this whole procedure to generate another 12656 mocks, with different shape noise, that we used to train the neural density estimators used for the likelihood-free inference. In total, therefore, we produced 25312 pseudo-independent mocks.

The process of creating mock datasets involves a number of unconstrained parameters, including four multiplicative shear biases, four redshift distributions, and the intrinsic alignment parameters. When generating each of the 25312 pseudo-independent mocks, we select one of these parameters randomly from their respective priors (as detailed in Table I). For the redshift distributions, for each mock we pick at random one of the multiple realisations provided by the HYPERRANK methodology [24] using photometric redshift data $x_{\rm phot}$; we then use it as a $\bar{n}(s)$. These realisations encompass the redshift calibration uncertainties. In Fig. 3, we present for each tomographic bin a few of the realisations used in this study. Finally, we used the 100 independent `CosmoGridV1` full-sky realisations to generate two sets (with and without baryonic feedback effects) of 400 independent DES Y3 shear mock maps.

III. SUMMARY STATISTICS

We use different Gaussian and non-Gaussian weak lensing summary statistics in this work. All the summary statistics are applied to weak lensing mass maps; as a first step, therefore, we create the weak lensing mass maps starting from the shear maps. This is achieved by using a full-sky generalisation of the [61] algorithm [57]. This produces noisy weak lensing mass maps in the form of HEALPix maps with a resolution of NSIDE = 512 (corresponding to a pixel size of ≈ 6.9 arcminutes). This procedure is repeated for all four tomographic bins of our catalogue. During the creation of the mass maps, we further applied a cut at $\ell_{\rm max} = 1024$. The maps at NSIDE = 512 formally have non-zero support up to $\ell = 1535$; most of their power, however, is suppressed above $\ell \sim 1000$ because of the pixel window function. We chose to incorporate this particular cut when we were constructing the pipeline as we were assuming then that we would need to remove these scales due to potential baryonic contamination; we did not revisit this choice after the scale cut test presented in §V, as it would have required us to redo the creation of the mocks and measurements.

The summary statistics considered in this work are: 1) second and third moments; 2) wavelet phase harmonics; 3) the scattering transform. The statistics are applied to smoothed versions of the weak lensing maps, with the type of smoothing depending on the statistic: moments use top hat filters, while wavelet phase harmonics and the scattering transform use wavelet filters [22, 73, 105]. In all cases we smooth the maps using filters with different sizes. More details and relevant equations are presented below.

A. Second and Third moments

The first statistics considered are second and third moments of the weak lensing mass maps [19, 39, 41, 83, 84, 106–108]. While second moments are a Gaussian statistic, third moments probe additional non-Gaussian information of the field. Second and third moments of the DES Y3 weak lensing mass maps have been recently used in [41] to infer cosmology; here we adopt that paper's implementation of the moments estimator.

We first smooth the maps using a top-hat filter with different smoothing scales. In practice, this is achieved by multiplying the coefficients of the harmonic decompositions of the weak lensing mass maps by

$$W_\ell(\theta_0) = \frac{P_{\ell-1}(\cos(\theta_0)) - P_{\ell+1}(\cos(\theta_0))}{(2\ell+1)(1-\cos(\theta_0))}, \tag{5}$$

where P_ℓ is the Legendre polynomial of order ℓ, θ_0 is the smoothing scale, and ℓ is the multipole. We consider eight smoothing scales equally (logarithmically) spaced from 8.2 to 221 arcmin, and we denote the smoothed lensing mass map of tomographic bin i by $\kappa^i_{\theta_0,p}$. We estimate the second and third moments as follows:

$$\langle\hat{\kappa}^2_{\theta_0}\rangle(i,j) = {\rm Avg}_p\left(\kappa^i_{\theta_0,p}\,\kappa^j_{\theta_0,p}\right) \tag{6}$$

$$\langle\hat{\kappa}^3_{\theta_0}\rangle(i,j,k) = {\rm Avg}_p\left(\kappa^i_{\theta_0,p}\,\kappa^j_{\theta_0,p}\,\kappa^k_{\theta_0,p}\right). \tag{7}$$

Here i, j, k refer to different tomographic bins; all combinations of tomographic bins are considered (ten independent combinations for second moments and 20 for third moments). The average is over all pixels p on the full sky (i.e. $1 \le p \le N_{\rm tot}$), including regions outside the footprint, since the Kaiser-Squires conversion, and the subsequent smoothing, transfers some of the signal from inside to outside the DES footprint.

We can only estimate *noisy* realisations of the weak lensing mass maps: $\kappa_{\rm obs} = \kappa + \kappa_{\rm N}$. Any statistic measured with data will include noise contributions [106]. When comparing measurements to analytical predictions, noise-only terms are normally subtracted to ease the comparison. If the noise-only terms are estimated from the data (via, for example, random rotation of the ellipticity measurements), subtracting the noise terms can increase the measurement uncertainties, because the noise terms estimates are affected by shot noise. While it would be possible to have multiple estimates of the noise terms for every map to reduce the shot noise contribution, we simply chose to not subtract these terms, except in a few particular cases.

As for moments, we decided only to subtract the following specific noise terms from our third moments estimator:

$$\langle \hat{\kappa}^3_{\theta_0} \rangle = \langle \hat{\kappa}^3_{\theta_0,\mathrm{obs}} \rangle - \langle \hat{\kappa}_{\theta_0,\mathrm{obs}} \hat{\kappa}^2_{\theta_0,\mathrm{N}} \rangle; \tag{8}$$

For third moments, we subtracted noise-signal third moments of the form $\langle \hat{\kappa}_{\theta_0,\mathrm{obs}} \hat{\kappa}^2_{\theta_0,\mathrm{N}} \rangle$. These terms are strictly non-zero because of spurious noise-signal correlations arising from source clustering; [42] found that subtracting these terms reduces the impact of source clustering (and hence potential biases in the analysis if the source clustering is mismodelled in simulations). Other terms ($\langle \hat{\kappa^2}_{\theta_0,\mathrm{obs}} \hat{\kappa}_{\theta_0,\mathrm{N}} \rangle$ and $\langle \hat{\kappa}^3_{\theta_0,\mathrm{N}} \rangle$) were not subtracted as they average to zero even in presence of source clustering [42].

B. Wavelet Phase Harmonics

Wavelet phase harmonics [5, 75] are the second moments of smoothed weak lensing mass maps that have undergone a non-linear transformation. The fields are first smoothed using a directional, multi-scale wavelet transform [22, 73, 105]; the wavelets have the advantage of being localised both in Fourier and real space, contrary to the top-hat filters used in this work for the second and third moments, which are local only in real space. Moreover, we adopt 'directional' wavelets, instead of using an isotropic filter.

We use the package PYWPH [3] to smooth our maps. The package works on a two-dimensional projection rather than on a sphere. Therefore, we first cut multiple square patches of roughly 14.6 degrees of side covering the full DES footprint. For this we use a gnomonic projection (as implemented in the HEALPix GNOMVIEW function), converting our patches to a 128x128 pixelated grid with a pixel scale of 6.8 arcminutes. Due to projection effects, the same portion of a map might appear in multiple projected patches; we mask pixels accordingly to avoid double-counting. Note that both simulated and real data maps undergo the same projection process.

We then smooth the projected patches using 'bump steerable wavelets'. Begin in Fourier space, where we define the wavelet

$$\hat{\psi}(\vec{k}) = \begin{cases} 0.7309 \exp\left(\frac{-(|\vec{k}|-\xi_0)^2}{\xi_0^2-(|\vec{k}|-\xi_0)^2}\right)\cos^2(\arg(\vec{k})) \\ \qquad \text{if } 0 < |\vec{k}| < 2\xi_0 \text{ and } k_x \geq 0, \\ 0 \quad \text{otherwise.} \end{cases} \tag{9}$$

Here $\vec{k} = (k_x, k_y)$ is the two-dimensional Fourier wavenumber, while ξ_0 denotes the central frequency of the wavelet (the full vector is $\vec{\xi_0} = (\xi_0, 0)$) and is set to $\xi_0 = 0.85\pi$ following [76]; the prefactor and the power of the cosine function corresponds to $L = 3$ in their notation. Note that $\hat{\psi}$ has finite width (i.e. 'compact support') in Fourier space. The real space Fourier transform ψ of this is then our 'mother' wavelet, from which other wavelets can be obtained by dilating and rotating:

$$\psi_{n,\ell}\left(\vec{\theta}\right) = 2^{-2n}\psi\left(2^{-n}\,\mathrm{Rot}_{-\ell}\,\vec{\theta}\right). \tag{10}$$

Here Rot_ℓ denotes rotation by an angle $\pi\ell/L$; we consider $L = 3$ (so that ℓ can be 0,1,2), corresponding to three possible orientations of the steerable wavelet. [4] The number n specifies an oscillation of the order of 2^{n+1} pixels; as we are using patches of 128x128 pixels, n runs from 0 to 5. This choice of spacing between different filter sizes follows the standard implementation of [5]; for simplicity, and in order to keep our data vector size reasonably small, we chose to not explore a thinner spacing. Note that the wavelet is real in Fourier space and is complex in real space.

The wavelet transform of a field is the convolution of the field with $\psi_{n,\ell}$ (for arbitrary n and ℓ). For the wavelet transform of the convergence map in tomographic bin i we write:

$$\kappa^i_{n,\ell}(\vec{\theta}) \equiv \left(\kappa^i * \psi_{n,\ell}\right)(\vec{\theta}). \tag{11}$$

Its Fourier transform for each (n, ℓ) has central frequency $\vec{\xi} = 2^{-n}\,\mathrm{Rot}_\ell\,\vec{\xi_0}$ and has finite width, and thus each convolution is a local filtering in Fourier space. As shown in Fig. 2 of [5], it can identify both peaks and anisotropic filaments of different orientations. The full wavelet transform spans all of Fourier space. In addition, it has the desirable feature of being well localised in both real and Fourier space.

Following [5], we apply a non-linear operation to the smoothed fields. The non-linear operation used is called 'phase acceleration'; this operation modifies the Fourier spectrum of the smoothed field, without modifying the spatial localisation of its features. As it is a non-linear operation, it allows us to access the non-Gaussian features of the field using second moments. Modifying the spectrum of the field, on the other hand, allows us to capture interactions between fields smoothed with different filters (and therefore different scales) that would otherwise have minimum overlapping support in Fourier space.

The smoothed and accelerated field will be called the wavelet phase harmonic. The 'phase harmonic of order q' is defined to be

$$\mathrm{PH}(re^{\mathrm{i}\theta}, q) \equiv re^{\mathrm{i}q\theta}, \tag{12}$$

where r is the modulus of the field and θ its phase. This function leaves its input unaltered for $q = 1$, and takes its modulus for $q = 0$. The absolute value of the field has

[3] https://github.com/bregaldo/pywph

[4] Note that this ℓ does not indicate the multipole of the spherical harmonic decomposition, as it is done in other sections of this paper, but rather the rotation index. We kept this notation in this section (and in the next one) to be consistent with the WPH literature.

been shown to be a useful non-linear operation, with the desirable property that it does not amplify noise. We consider only $q = 0$ or $q = 1$; although q can reasonably assume other values [5], we found these other statistics did not significantly improve the constraints.

Once the fields have been transformed, we can build statistics that are second order in the input field, of the form:

$$\mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(\mathrm{PH}(\kappa^i_{n_1,\ell+\Delta\ell}, q_1)\;\mathrm{PH}(\kappa^j_{n_2,\ell}, q_2)\right). \tag{13}$$

As before, we average over all pixels. We also average over all values of the rotation index ℓ (i.e. $0 \le \ell < L$); note that this makes sense even when $\Delta\ell \neq 0$ as the rotation indices can simply 'wrap around'. These statistics are therefore functions of scales (n_1, n_2), rotation index offset ($\Delta\ell$), phase harmonic orders (q_1, q_2), and tomographic bins (i, j). The statistics used in this work are:

$$S00(i,j,n) = \mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(|\kappa^i_{n,\ell}|\,|\kappa^j_{n,\ell}|\right) \tag{14}$$

$$S11(i,j,n) \equiv \mathrm{WPHG}(i,j,n) = \mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(\kappa^i_{n,\ell}\,\kappa^j_{n,\ell}\right) \tag{15}$$

$$S01(i,j,n) = \mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(|\kappa^i_{n,\ell}|\;\kappa^j_{n,\ell}\right) \tag{16}$$

$$C01\delta\ell0(i,j,n_1,n_2) = \mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(|\kappa^i_{n_1,\ell}|\;\kappa^j_{n_2,\ell}\right)\;\text{for } n_1 < n_2 \tag{17}$$

$$C01\delta\ell1(i,j,n_1,n_2) = \mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(|\kappa^i_{n_1,\ell+1}|\;\kappa^j_{n_2,\ell}\right)\;\text{for } n_1 < n_2. \tag{18}$$

Here i, j vary over the four tomographic bins, whereas n (or n_1 and n_2) varies over the possible wavelets under consideration. Following [5] we use 'S' for the statistics with $n_1 = n_2 = n$ ($S00$, $S11$, and $S01$) and 'C' for the statistics with $n_1 < n_2$ ($C01\delta\ell0$ and $C01\delta\ell1$) that capture correlations at different wavelet scales.

The statistics probe non-Gaussian features of the field (with the exception of $S11$, which is Gaussian in that it is equivalent to the power spectrum of κ; for this reason we refer to it as 'WPHG'). One advantage of the WPHs over conventional moments is that they are always 'second-order' in the input field, which makes them more robust against additive noise [5]. Additional statistics using more combinations of WPHs could have been considered, as in [5]; however, for computational reasons we restrict ourselves to the summary statistics listed (having checked that they capture nearly all the information given the noise levels in our data). In total, we have 60 components for $S11$ (ten independent tomographic bin pairs and six scales), 96 components for $S00$ and $S01$ each (16 tomographic bin pairs and six scales), and 240 components for $C01\delta\ell0$ and $C01\delta\ell1$ each (16 tomographic bin pairs and 15 scale pairs).

As in the case of moments, we subtract some specific WPH moments of noise-only maps from our estimators. In particular, for WPH $S01$, $C01\delta\ell0$, and $C01\delta\ell1$ we subtract a term involving one noise-only map and the observed noisy convergence map. We empirically found these statistics to be the ones mostly affected by source clustering, and this subtraction to be the best way to minimise source clustering effects.

C. Scattering Transform

The scattering transform [18, 21, 74, 102, 103] is in concept similar to the WPHs introduced above. The idea is to smooth the field using the directional, multi-scale wavelet transform, followed by a modulus operation on the field. This pair of operations can then be reapplied several times; we finish with an overall average over the sky. This yields a hierarchy of scattering transform coefficients ST_m, where m is the number of smoothing and modulus operations applied. This work uses scattering coefficients of order $m = 1, 2$. Given a directional multi-scale wavelet $\psi_{n,\ell}$ and the convergence map κ^i of tomographic bin i, we obtain:

$$\mathrm{ST}_1(i,n) = \mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(|\kappa^i * \psi_{n,\ell}|\right) \tag{19}$$

$$\mathrm{ST}_2(i,n_1,n_2,\ell') = \mathrm{Avg}_p\,\mathrm{Avg}_\ell\left(|\,|\kappa^i * \psi_{n_1,\ell}| * \psi_{n_2,\ell'-\ell}|\right)\;\text{for } n_1 \le n_2. \tag{20}$$

The average over ℓ in ST_2 makes this summary statistic invariant to rotation, while preserving morphological information. The ST_1 coefficients are qualitatively similar to the power spectrum amplitudes, weighted by a window function, but while the power spectrum uses the L^2 norm of the convolved field, the scattering transform uses the L^1 norm. The ST_2 coefficients probe more non-Gaussian information stored in the field, providing the co-occurrence information at the scales n_1 and n_2 and capturing interferences of the field between features selected with two successive wavelets.

The scattering transform coefficients are 'first-order' in the input field. To enable the computation of scattering transform coefficients including pairs of maps of different tomographic bins, we follow [113] and introduce the 'cross-maps' $\kappa^{ij}(\theta,\phi)$

$$\kappa^{ij}(\theta,\phi) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} \hat{\kappa}^i_{\ell m}\hat{\kappa}^j_{\ell m} Y_{\ell m}(\theta,\phi), \tag{21}$$

where i and j (with $i > j$) denote two different tomographic bins. We then compute the scattering coefficients ST_1 and ST_2 of the cross-maps. In total, we consider 60 coefficients for ST_1 (six scales, ten independent tomographic bins), and 630 for ST_2 (21 scale combinations, ten tomographic bins, and three different orientations).

The ST is similar to the WPHs, but with a few differences. First, the scattering transform stays 'first-order' in the observed field, whereas the WPHs are always 'second-order'. This means that the ST is less susceptible to noise than the WPHs. Second, in the WPHs there is

a natural definition of cross-correlation between different fields; this is not the case for the ST (it is for this reason that we introduced the cross-maps so as to account for cross-correlations between different tomographic bins). As we will see in §VI, this has an impact on the constraints, as the ST deals with cross-correlations less efficiently. This also applies to cross-correlations between different scales for the non-Gaussian features: the WPHs use the cross correlations in combination with the phase acceleration as a non-linear operation to couple scales (e.g. the WPH C01 coefficients); an analogous statistic cannot be defined for the ST.

ST and WPH are often compared to machine learning methods as they were designed to emulate information capture in the manner of a convolutional neural network (CNN), without the need for training data. This is quite evident especially for the ST coefficients: the smoothing of the field is equivalent to the CNN kernel convolution, the modulus operation is equivalent to the CNN ReLU layer, the average is equivalent to the CNN 'pooling', and the hierarchy of coefficients is equivalent to the CNN's multiple layers. The analogy, however, stops here: for the ST and WPH, since there is no training, we have full control over the kernels (i.e. the wavelets), or on all the details of the summary statistics (i.e. the order of the phase acceleration for the WPH, or how different tomographic bins are combined). This is different to CNNs, which are commonly referred to as 'black boxes' because of the difficulties associated with comprehending the features they learn and the significance of the numerous parameters acquired during training.

D. Multipole support, covariance, and signal-to-noise of the summary statistics

The statistics considered here implement different filters to smooth the convergence field, so it is instructive first to look at the support in multipole space covered by the smoothed maps. This is shown in Fig. 4, which plots (for the first and last tomographic bin convergence maps) the power spectra of the smoothed maps, where this smoothing is done using top-hat and directional wavelet filters of different sizes. To compare roughly the two types of filter, the top hat filter radii θ have been chosen to be half the FWHM of the wavelet filters. Fig. 4 shows that the maps smoothed by the two sets of filters peak roughly at the same point in multipole space, but that the top hat filters are much less localised. This is expected as wavelets are designed to better isolate scales, both in real and in multipole space.

The statistics considered here are also in part covarying, i.e. they probe similar information. Therefore it is instructive to construct the correlation matrix of the data vector; this can be done starting from the 400 mock measurement of the `CosmoGridV1` simulations and by computing the covariance matrix:

$$\hat{C} = \frac{1}{\mathrm{N_s}} \sum_{i=1}^{\mathrm{N_s}} (\hat{d}_i - \hat{d})(\hat{d}_i - \hat{d})^T, \tag{22}$$

where $\mathrm{N_s}$ is the number of simulations, $\hat{d}_i$ the data vector measured in the i-th simulation, and $\hat{d}$ the sample mean. The elements of the correlation matrix $\widehat{\mathrm{Corr}}$ can be obtained as

$$\widehat{\mathrm{Corr}}_{i,j} = \frac{\hat{\mathrm{C}}_{i,j}}{\sqrt{\hat{\mathrm{C}}_{i,i}\hat{\mathrm{C}}_{j,j}}}. \tag{23}$$

This is illustrated in Fig. 5, which shows the correlation matrix of the different statistics as a function of scales. For the sake of simplicity, we considered only the part of the data vector including the fourth tomographic bin. We make three remarks:

- Second and third moments blocks are much more correlated than those of the scattering transform and the WPH. This is a consequence of the smoothing filter adopted: wavelet filters are significantly better at isolating scales, and this makes the correlation matrix more diagonal.

- Gaussian statistics (second moments and WPHG) are highly correlated, as expected. They are also highly correlated with WPH S00 and the scattering coefficients ST_1 and ST_2. The latter are probing both Gaussian and non-Gaussian features of the field, although this figure suggests they weigh Gaussian features more.

- Third moments and WPH $S01$ are not very correlated with their Gaussian counterparts (a fact also exacerbated by shape noise), but they are mildly correlated with each other. This suggests WPH $S01$ is in part probing the bispectrum of the field. Although not shown in the figure, we report that WPH $C01\delta\ell0$ and $C01\delta\ell1$ behave similarly to WPH $S01$.

We report the signal-to-noise ratio (SN) of the different statistics in Table II. We note that this SN is computed for the part of the measurements that only uses one tomographic bin. The Gaussian statistics considered in this work have significantly higher signal-to-noise compared to third moments or WPH $S01$. On the other hand, WPH $S00$, ST_1 and ST_2 have significance similar to Gaussian statistics, as they are also probing Gaussian information of the field. Among the purely non-Gaussian statistics, we note that WPH $S01$ has a higher signal-to-noise ratio than that of the third moment. This is due to the former statistic being only 'second-order' in the input field, which makes it less affected by noise.

Last, we show in Fig. 6 some of the statistics as measured in `CosmoGridV1` simulations at the fiducial cosmology.

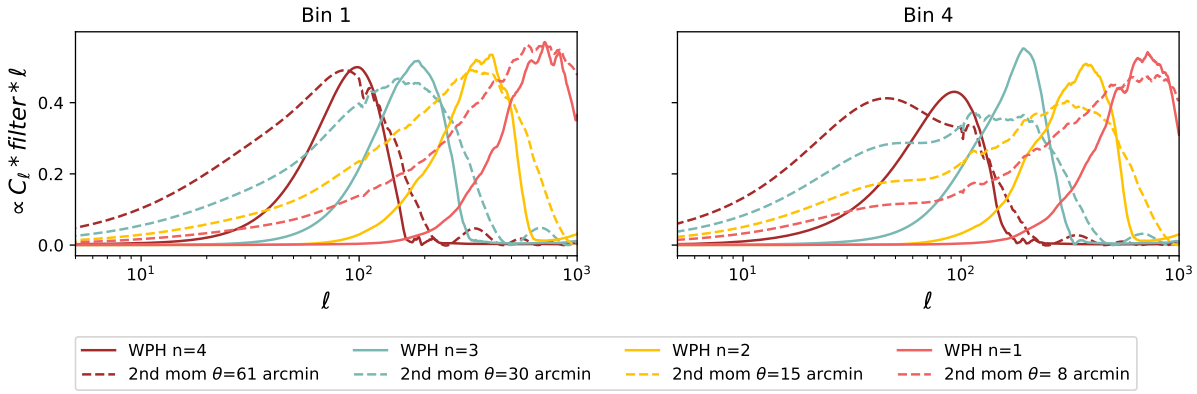

Figure 4. Power spectra (C_ℓ) of convergence maps of the first (left) and last (right) tomographic bins; the maps have been smoothed by directional wavelet (solid line) and top hat (dashed line) filters of different sizes. For plotting purposes the measured power spectra have been smoothed with a Savitzky–Golay filter.

Table II. Salient properties of the summary statistics. The second column denotes whether it carries Gaussian (G) or non-Gaussian (NG) information. The third column refers to the order of the field κ. The fourth column is the number of components of the datavector across scales and tomographic bins. The further columns show the signal-to-noise ratio (SN) of the measurements in the `CosmoGridV1` simulations for each tomographic bin. We note that this is not the total SN of the full measurement, but only the SN of the measurement for one tomographic bin.

	G/NG	**Order**	**Length of Datavector**	**Bin 1 SN**	**Bin 2 SN**	**Bin 3 SN**	**Bin 4 SN**
2nd moments	G	2	160	3.4	7.8	16.1	15.2
3rd moments	NG	3	512	0.8	0.9	1.7	1.3
WPH S11 (WPHG)	G	2	120	3.1	7.4	15.6	14.4
WPH S00	NG	2	96	2.8	6.9	14.9	13.6
WPH S01	NG	2	480	0.7	1.5	2.9	2.4
ST1	NG	1	60	3.3	7.8	15.3	15.3
ST2	NG	1	630	3.1	7.3	15.3	15.3

E. Data Compression

Data compression is paramount in the likelihood-free inference framework, as for a fixed number of simulated mocks the density estimation is more efficient when the dimensionality of the data vector is low [56]. Different compression methods exist (e.g. PCA-based compression, Zürcher et al. 112; MOPED, Heavens et al. 50; neural compression, Jeffrey et al. 56). Notably, a poor compression scheme could result in less informative summaries, but it would not produce biased results. For this work we follow Jeffrey et al. 56 and use a neural compression scheme to compress the summary statistics to the same dimension as the parameters θ in which we are interested (but see Appendix B for a comparison with the MOPED compression). In particular, given a summary statistic $\mathbf{d}$, we compress it using $\mathbf{t} = F_\phi(\mathbf{d})$, and we approximate F_ϕ by a neural network. We determine the neural network parameter ϕ by minimising a Mean Squared Error (MSE) loss function using the first half (12656) of our pseudo-independent mocks. The architecture used for the network and the number of parameters are summarised in Table III. Since in this work we consider multiple summary statistics and their combinations, we chose to compress summary statistics individually and to combine their compressed versions (i.e. stack the data vectors) later on during the likelihood-free inference process. In particular, we individually compress second moments, third moments, **WPHG**, WPH $S00$, $\mathbf{ST}_1$, and $\mathbf{ST}_2$. The only exception is for WPH $S01$ and WPH $C01$, which are compressed together. We compress the data vectors using all the parameters, one at a time. Examples of compression are shown in Fig. 7 for second moments and WPH S01+C01, against the parameters Ω_m and S_8. Generally, the tighter the scatter, the better the given statistic is at constraining that parameter. For second moments, the compressed statistics trace fairly well the parameter against which they have been compressed; on the other hand, the WPH S01+C01 case shows a poor sensitivity to Ω_m. The compression is not expected to be 'unbiased': as it can be seen from Fig. 7, the compressed statistics do not recover the true value of the simulations (the red line in the plot), even in the best case (S_8 for second moments). This is not a problem for the inference; as we consistently compress both the data vectors measured in simulations and the data, the final posterior will be unbiased.

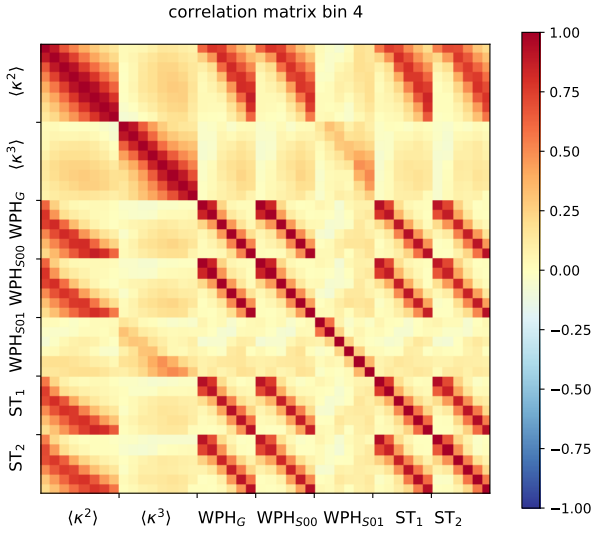

Figure 5. Correlation matrix for some of the statistics considered in this work, computed from the `CosmoGridV1` simulations at the fiducial cosmology. We consider only the fourth tomographic bin.

Table III. Neural Network Layers and number of parameters used for the compression of the summary statistics.

Layer (type)	Output Shape	Number of Parameters
Dense	900	900*(length DV+1)
LeakyReLU	900	0
Dense	800	720800
LeakyReLU	800	0
Dense	100	80100
ReLU	100	0
Dense	100	10100
ReLU	100	0
Dense	1	101

IV. LIKELIHOOD-FREE INFERENCE

In likelihood-free inference (also known as simulation-based inference), the likelihood $p(d|\theta)$ is not assumed to have a closed form; rather, it is reconstructed from simulated mock data as part of the inference pipeline. Here is a summary of the procedure used to infer the posterior distribution of the parameters; a more detailed description is provided in [56].

In our implementation, the parameter inference task is posed as a density estimation problem. Let us assume we have a set of mock noisy data vectors d and simulation parameters θ forming a cloud of points in $\{d, \theta\}$ space. We then estimate the conditional distribution $p(d|\theta)$ with an ensemble of neural density estimators (NDEs): specifically, we use both Gaussian Mixture Density Networks (MDN; [13]) and Masked Autoregressive Flows (MAF; [80]). We used two MDNs with two and three Gaussian components respectively, each with two dense hidden layers with 30 neurons per layer, and we used two MAFs with two and three MADE (Masked Autoencoders for Distribution Estimation, [44]) layers

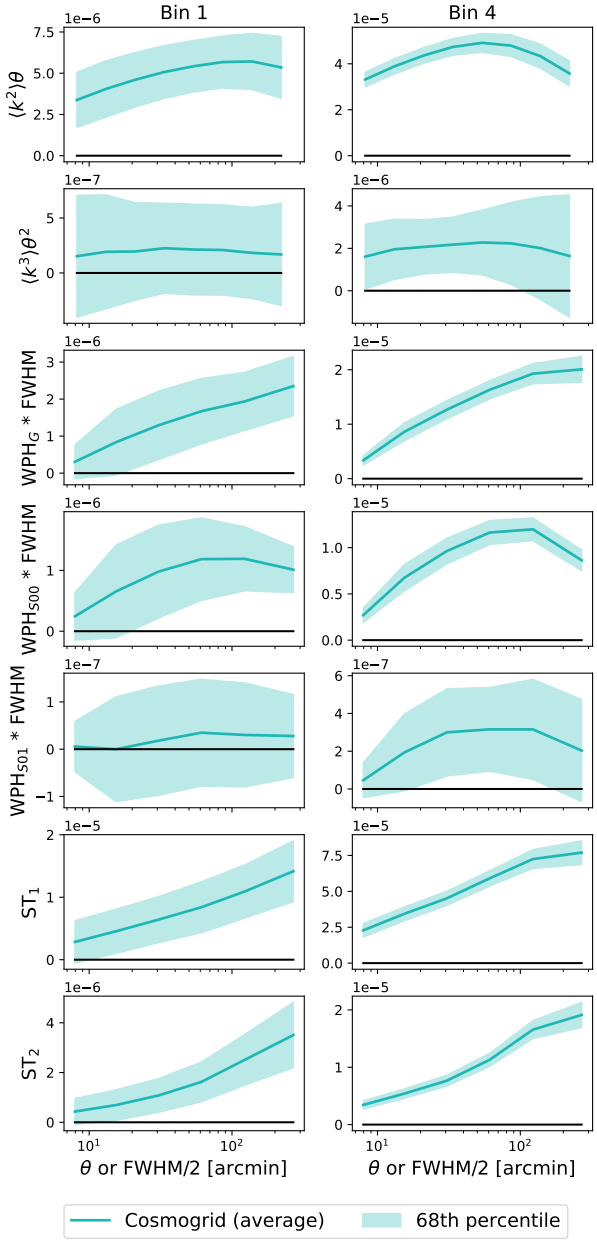

Figure 6. Some of the Gaussian and non-Gaussian statistics considered in this work, as measured in simulations. The left (resp. right) column shows the statistics measured only using the noisy convergence map of the first (resp. fourth) tomographic bin. The calculation was done for multiple `CosmoGridV1` simulations at the fiducial cosmology; solid lines are the resulting average and the bands are 68th percentiles. For statistics involving filters with different sizes j_1, j_2, we considered $j_1 = j_2$.

respectively, each with two dense hidden layers with 50 neurons per layer. For each of our neural density estimation methods, MDN and MAF, the network was trained to give an estimate $q(d|\theta;\phi)$ of the target distribution $p(d|\theta)$ (i.e. $p(d|\theta) \approx q(d|\theta;\phi)$); here ϕ are the parameters of the network, determined by minimising a loss function $U(\phi) = -\sum_{n=1}^{N} \log q(d_n|\theta_n;\phi)$ over the N forward-modelled

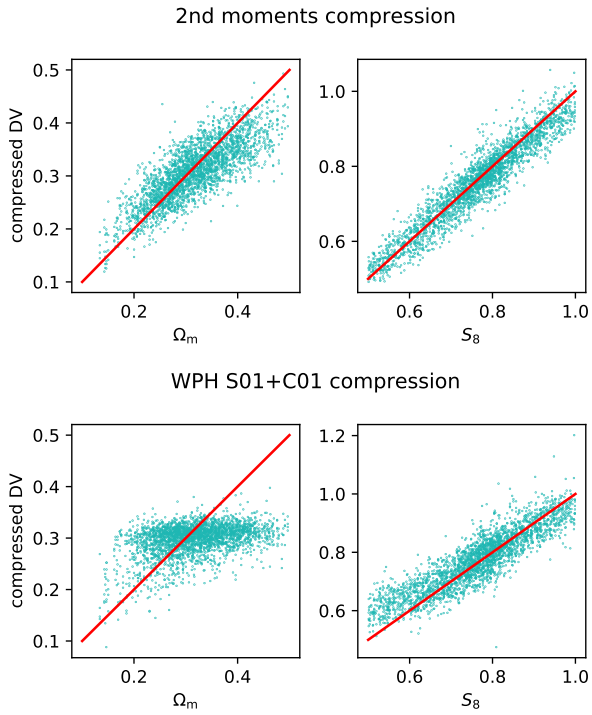

Figure 7. Example of compressed statistics using second moments (top) and PWH S01+C01 (bottom) as summary statistics, and Ω_m and S_8 for target parameters for the loss function. The y-axis is the compressed statistic, while the x-axis is the true value of the parameter. Each point represents one input measurement. The red line serves to guide the eye and indicates an unbiased compression. Generally, the tighter the scatter, the better the given statistic is at constraining that parameter; on the contrary, a broader and biased compression indicates poor sensitivity to that parameter (e.g. WPH S01+C0 for Ω). A biased compression does not imply a biased inference, as we compress the simulated measurements and the data in the same way.

mock data d_n. This loss corresponds to minimizing the Kullback-Leibler divergence, a measure of difference or change going from the estimate q to the target $p(d|\theta)$. To perform the density estimation and the training we used the publicly available package pyDELFI [6].

The final density estimation is a stack of the ensemble estimates, weighted by the loss evaluated during training. Once the target distribution $p(d|\theta)$ has been estimated, we evaluate it at the observed data $d = d_{\rm obs}$ to obtain the likelihood. For completeness, we show in Appendix C the posteriors obtained by each NDE and how they differ from the stacked posterior.

Using NDEs to infer the *likelihood* surface rather than the posterior has one main advantage: as long as the parameters varied in the simulations are taken into account during the training process, the fact that the parameter space is not sampled uniformly does not translate into an effective prior on our final constraints, i.e. it does not produce tighter posteriors [6]. This means that after we trained the NDEs and learned the likelihood surface, we can use a different prior during the inference when estimating our posteriors (see Table I for the priors used in the analysis). Of course, in the regions of the parameter space where we only have a few simulations, the estimation of the likelihood surface will be noisier and the likelihood less accurate; this is why the Gower St simulations have been run in active-learning mode for Ω and S_8, to increase the accuracy of the likelihood estimation in the region covered by the data posterior.

For practical reasons, due to our limited number of mocks, it is not possible to reliably estimate the likelihood surface taking into account all the parameters varied in the simulations. As we are mostly interested in the constraints on Ω_m, S_8, w, and $A_{\rm IA}$, the main density estimation was carried out using the parameters $\theta = [\Omega_m, S_8, w, A_{\rm IA}]$ and the associated compressed data vectors. This means that the other parameters are effectively marginalised over; this time, however, since we are not explicitly taking into account their dependence during the training of the NDEs, the parameter distribution *does* matter. This is explained via *marginal posterior density* estimation in [58]; we can therefore assume their marginalisation follows the prior distribution used to sample these parameters when generating the mocks as reported in Table I.

To train the NDEs we used the compressed data vectors and mocks that were not used to train the compression algorithm (i.e. the last 12656 pseudo-independent mocks). Whenever we combine different summary statistics, we stack the individual compressed data vectors together. We restrict the density estimation procedure to our eventual prior range (Table I). The final posteriors are then obtained through Markov chain Monte Carlo (MCMC) sampling of the likelihood, assuming the priors listed in Table I. The MCMC sampling is performed using the public software package `EMCEE` [37], an affine-invariant ensemble sampler for MCMC.

To test that the confidence levels obtained through the likelihood-free-inference are not misestimated, we perform an empirical coverage test. We first select a subset (125) of the full-sky Gower St simulations uniformly spanning the $\Omega_m - S_8 - w$ space. We do this by uniformly dividing each dimension into 5 parts, so as to partition the three-dimensional space into 5x5x5 cuboids, and by selecting only one simulation per cuboid. For this test, we excluded the outermost regions of our parameter space close to the edge of the priors, where we know we only have a few simulations and the likelihood estimation might be uncertain: in particular, we only selected simulations in the range $\Omega_m \in [0.2, 0.4]$, $\S_8 \in [0.6, 0.9]$, and $w \in [-1, -0.5]$. For each of the full-sky simulations, we choose four non-overlapping DES Y3 mocks (picked at random from the different noise realisations), for a total of 500 mocks. We re-train our compression algorithm and NDEs excluding these mocks; then, we obtain posteriors for each of them and check the confidence regions that cover the true values of Ω_m and S_8. Finally, we report in Fig. 8 the fraction of posteriors encompassing the true value at a given confidence level. A perfectly calibrated posterior would have an expected

coverage probability equal to the credibility level. Overconfident posteriors (i.e. tighter than they should be) would lie in the bottom right part of the plots; on the other hand, conservative posteriors (i.e. larger than they should be) would lie in the upper left part of the plot. The number of posteriors we ran limits the accuracy of this test; with 500 posteriors per summary statistics, we can determine if the posterior size is accurate at the ∼ 5 percent level. The statistics considered are consistent with a perfect calibrated posterior within the accuracy of the test. When all the posteriors are considered (lower panel of Fig. 8), the scatter reduces significantly, indicating no bias in the size of the posterior at the level of a few percent.

In Appendix D we provide further tests concerning the NDE likelihood estimates using `CosmoGridV1` simulations.

A. Comparison between approaches with theory-based models and Gaussian likelihood

We perform in this section a comparison between a) the cosmological constraints obtained using the LFI pipeline and b) a more standard approach in which we rely on a theoretical model for the observables and we assume the likelihood to be Gaussian. To this end, we use as a summary statistic the (pseudo) power spectrum, as implemented in [28]. Most of the summary statistics explored in this work do not have a theoretical model, except for the second and third moments [39, 41]; the code available to us to model moments, however, does not allow us to marginalise over the neutrino mass or over w. Moreover, we do not have a theoretical model for the covariance, which is, on the contrary, available for the pseudo power spectrum analysis. For these reasons we decided to use the power spectrum as a summary statistic for this comparison.

To perform the comparison, we analysed a theory data vector at a fiducial cosmology. As a minor caveat, we created (specifically for this test) mocks without source clustering (i.e. we assumed $b_g = 0$ and $F(p) = 1$ in Eq. 2), as this effect is not included in the theory model for the power spectrum; moreover, with source clustering the noise is slightly cosmology dependent, and this effect is not captured by the theory covariance implemented in [28]. Without source clustering, we note that the IA model reduces to a pure NLA model.

We then analysed the same noisy data vector using the theory model of [28], which is based on halofit [99]. We sampled the posteriors of our parameters using PolyChord [47, 48]; this is a nested sampler that uses slice sampling within the nested iso-likelihood contours. For the cosmological parameters, we varied the same parameters spanned by our mocks (see Table I), and, where possible, we assumed the same priors. For Ω_b, h_{100}, n_s, and neutrinos, we assumed flat priors, but we later importance-sampled the posterior to reflect the Gower St effective priors.

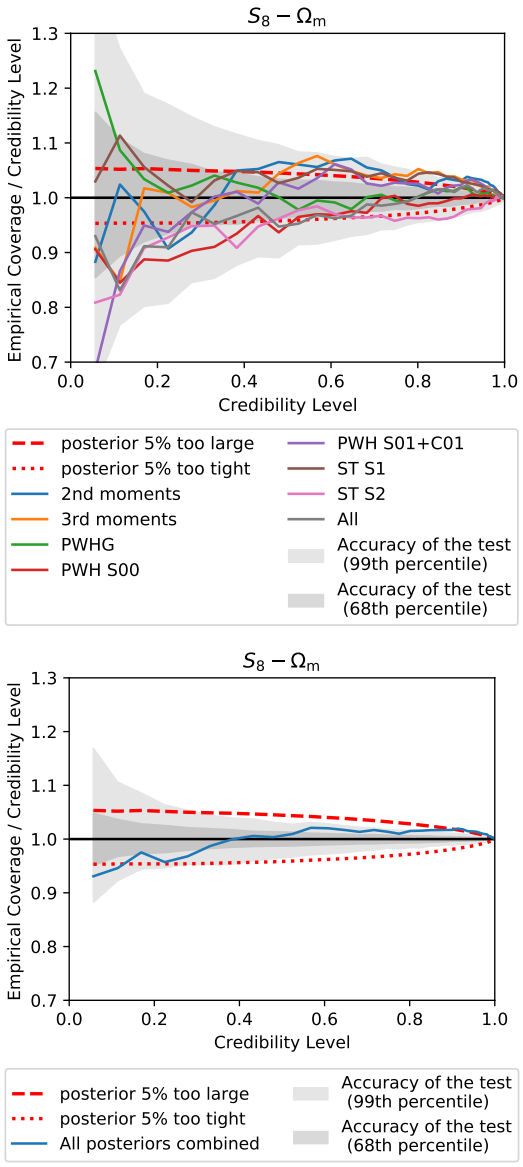

Figure 8. Expected coverage probability of the posteriors obtained using the LFI pipeline and different summary statistics with respect to the credibility level. The two red dashed/dotted lines indicate what the expected coverage probability would be if the posteriors were misestimated by 5 per cent. The grey shaded regions indicate the accuracy of the test given the limited number of posteriors (500) used here. The top panel shows the test for each of the summary statistics considered in this work and their combination; the bottom panel uses all the posteriors (500x8=4000) to test the size of the posteriors with a higher accuracy.

The posteriors for S_8 and Ω_m from the two pipelines are shown in Fig. 9, showing an excellent agreement. This agreement is not trivial: it relies on the validity of the Gaussian likelihood assumption for the power spectrum analysis, on the forward modelling of our simulations to be equivalent to the modelling used by the theory pipeline of [28], and on the dependence of the covariance on cosmological and nuisance parameters to be negligi-

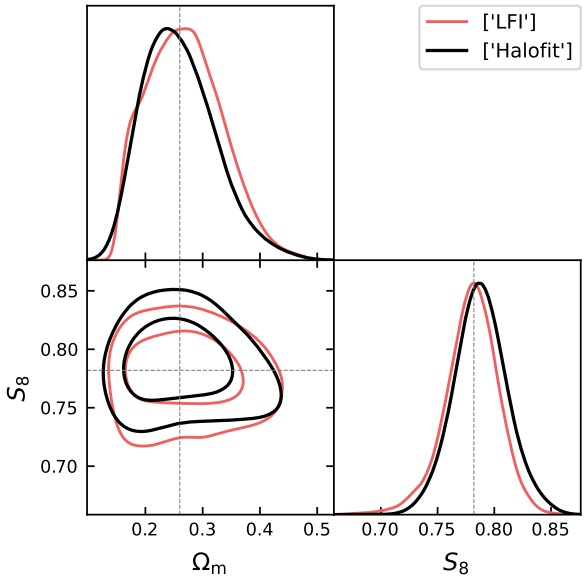

Figure 9. Posteriors for S_8 and Ω_m obtained by analysing the power spectrum of one of the Gower St simulations with two pipelines: the LFI pipeline described in this work and a different one that uses the theory model described in [28] and that assumes a Gaussian likelihood.

ble. In other words, a lack of agreement would not have invalidated our pipeline; rather, it would have challenged some of the main assumptions behind standard Gaussian likelihood analyses of Gaussian statistics such as found in [28]; [7]; [96]. The primary validation tests for assessing the accuracy of our posterior estimates include the empirical coverage test outlined in the preceding section, as well as the scale-cut tests and the end-to-end pipeline test discussed in the subsequent sections (§V and §VI).

V. SCALE CUTS

We determine in this section if we need to remove scales from our analysis because of a lack of modelling and/or potential systematic contamination. We test three main effects: 1) baryonic feedback processes; 2) additive biases due to PSF errors; 3) residual source clustering contamination. To anticipate the results of this section, we state here that we found all these effects to be negligible; therefore, our main analysis retains all the scales considered so far.

A. Impact of lack of modelling Baryonic feedback

The main limitation of our analysis is the lack of a proper model for baryonic feedback processes at small scales. The modelling of our observables relies on our ability to produce realistic mock catalogues; at small scales, this requires an ability to contaminate the mock catalogues with a variety of baryonic feedback models. Tools to create such contaminated catalogues exist; for example, baryonic correction models [9, 93] can adjust the particle positions in gravity-only simulations to mimic the impact of various baryonic processes on the density distribution. These models have been shown to be flexible enough to accurately replicate the 2-point and 3-point statistics of various hydrodynamical simulations. Unfortunately, the simulations we use for this project have not been post-processed with the baryonic correction model, forcing us to remove scales that can be potentially affected by baryons. This is also in line with the main DES Y3 strategy for weak lensing analyses (e.g. Amon et al. 7;Secco & Samuroff et al. 96; Gatti et al. 41, Zürcher et al. 112), which did not attempt to model baryonic processes but rather removed scales potentially affected by them.

Table IV. Bias of the parameter posteriors assessed by comparing the outcomes of an analysis performed on a simulation with baryonic feedback to that of a simulation without baryonic feedback. Biases for different summary statistics are reported in terms of the distance between the peaks of the posteriors in the $S_8 - \Omega_m$ plane. All the biases are smaller than 0.3σ (the maximum level of bias accepted by our analysis).

Summary Statistic(s)	**Contamination $S_8 - \Omega_m$**
2nd moments	0.01σ
WPHG	0.05σ
3rd moments	0.09σ
WPH S00	0.01σ
WPH S01+C01	0.04σ
WPH S00+S01+C01	0.11σ
ST1	0.03σ
ST2	0.03σ
ST1+ST2	0.06σ
2nd+3rd moments	0.03σ
2nd moments+WPH	0.03σ
2nd moments+ST1	0.05σ
2nd+3rd moments+ST+WPH	0.05σ

To determine which scales to remove, we use another set of public gravity-only simulations (`CosmoGridV1`) that have been post-processed with the baryonic correction model. For each full-sky simulation (with and without the baryonic correction model), we cut out four DES Y3 footprints and produce ten different noise realisations using our pipeline, totalling to two sets of 400 DES Y3 mock catalogues. The impact of the baryonic feedback model on the statistics used in this work is shown in Fig. 10.

The main effect of the baryonic model adopted is to suppress the values of the measured statistics, at all scales, with more dramatic effects in the first tomographic bin (first and second rows of Fig. 10). Statistics based on wavelets seem to be affected more by baryonic feedback than moments, as the latter rely on top-hat smoothing. This is, however, not a real problem, as it is due to the top hat filters being broader and skewed towards smaller multipoles / larger scales, not affected by baryons; this dilutes the baryonic contamination. For

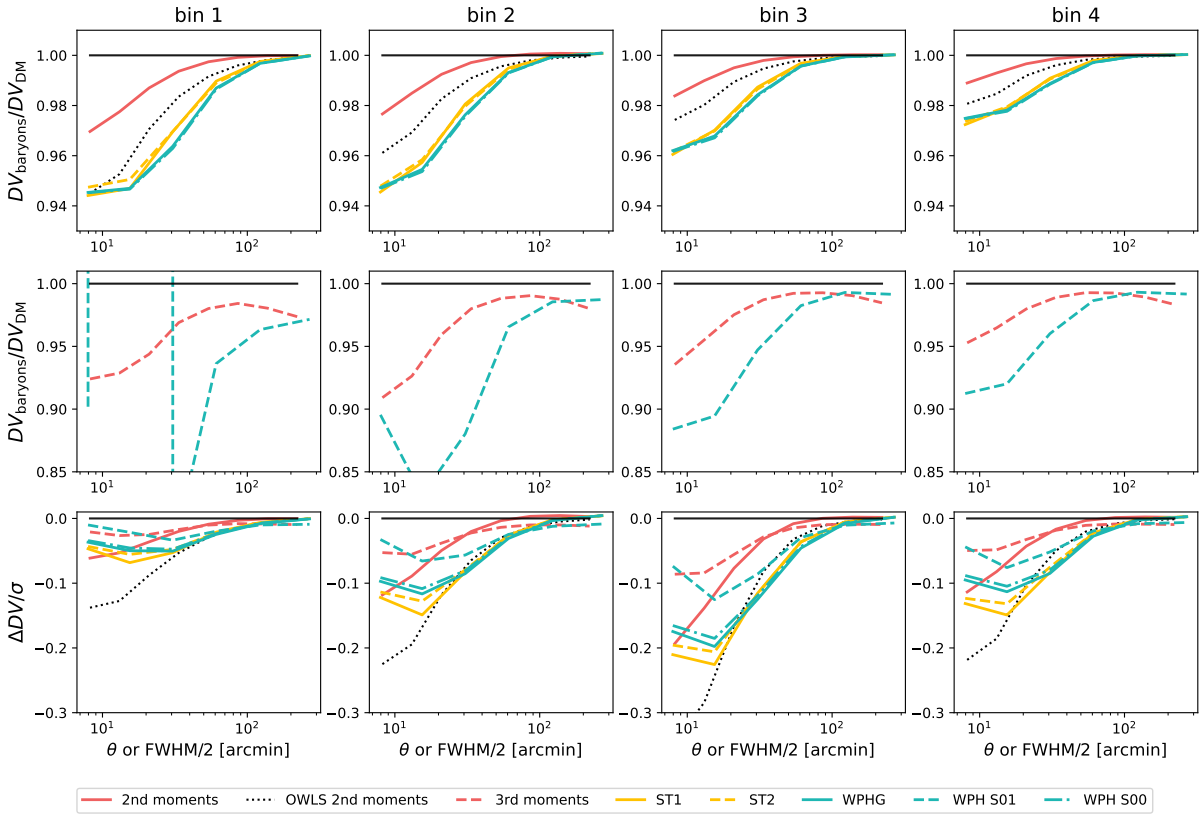

Figure 10. Impact of baryonic feedback effects on the summary statistics considered in this work. Each column shows the impact on the part of the summary statistics obtained using maps from a specific tomographic bin. The top two rows show the ratio between the data vector as computed in simulations with and without baryonic feedback; the bottom row shows the difference between the data vectors normalised by the square root of the diagonal of the covariance matrix. For the second moment, we also show the result for the OWLS-AGN simulations (dotted line as indicated in the legend). We note that in the first bin, the ratio between the WPH S01 data vector with and without baryonic contamination changes sign at small scales; this is because at those scales the amplitude of the data vector changes sign and it is close to zero.

practical purposes it is actually better to have filters with a more compact support, as this makes it easier to remove the part of the measurements affected by systematics.

The impact of baryons on non-Gaussian statistics can be qualitatively different from their Gaussian counterparts [9, 36]. In a first approximation a suppression of the underlying density field should translate into a suppression of N-point statistics that will be larger as the order of the statistics increases. Fig. 10 indeed shows a larger impact of baryons on the amplitude of the data vector for third moments compared to second moments. For the other non-Gaussian statistics included in this work, however, it is more difficult to apply this qualitative argument: ST and WPH are either linear or second order in the input field, and many of them are highly correlated with the Gaussian statistics. The impact of baryons on the amplitude of WPH S01 is significantly larger compared to WPHG, but for all the other non-Gaussian statistics, the suppression is basically the same as that of WPHG.

To determine which scales to remove from our analysis, we check that the posterior on the cosmological parameters obtained by analysing a data vector from the simulations with baryonic feedback is not substantially biased with respect to the posterior obtained from a data vector measured in simulations without baryons. We adopted the same criterion used by the main DES cosmological analysis (Amon et al. 7; Secco & Samuroff et al. 96; Abbott et al. 1). The criterion requires the peak of the marginalised two dimensional posterior of $\Omega_{\rm m}$ and $S_8 \equiv \sigma_8(\Omega_{\rm m}/0.3)^{0.5}$ obtained by analysing the contaminated data vector to be within 0.3σ of the values obtained with the uncontaminated one. We note that the baryonic model adopted by the main DES analyses to determine the scale cut follows the predictions from the OWLS 'AGN' simulations [91, 104]. The baryonic feedback of the `CosmoGridV1` simulations, however, is slightly milder then the OWLS model. This difference is illustrated in Fig. 10, where we also show the impact of the OWLS AGN feedback on second moments, computed following the method in Gatti et al. 41.

The level of contamination obtained using all the scales at our disposal is reported in Table IV, for a subset

of individual summary statistics and for (some) of their combinations. Fortunately, none of the summary statistics exceed our predefined criteria for contamination, which confirms the robustness of our analysis against potential baryonic feedback processes. While it is true that the baryonic model of the `CosmoGridV1` simulations is milder than the OWLS model, these numbers are safely smaller than 0.3σ. For second moments only, where we can compute the impact of the OWLS AGN feedback analytically, we also analysed a theory data vector 'contaminated' with the OWLS AGN feedback, finding only a 0.1σ shift with respect to dark-matter-only data vector. In hindsight, we realize that we could have generated maps with higher resolution, even beyond NSIDE = 512 (~ 7 arcmin). Such higher resolution would have allowed us to explore smaller scales, but it would have come with a considerable increase in computational cost, which we choose to defer to future research.

B. Impact of potential mismodelling of source clustering effects

Source clustering refers to the angular distribution of source galaxies being not uniform, but rather being modulated by clustering due to galaxies tracing the underlying density field [42, 64, 92, 95, 101]. This effect causes the galaxy number density to be correlated with the target lensing signal: since we expect a larger lensing signal along overdense lines-of-sight, we preferentially sample the shear field where its value is larger.

For estimators based on pixelized shear maps, this has two effects [42]:

- the average noise-free lensing signal is modulated by a different effective redshift distribution;
- the shape noise in every pixel is correlated with the lensing signal.

The first effect is generally small. The second effect can be large for non-Gaussian statistics whenever the estimators used involve correlation between the lensing signal and even moments of the noise (e.g. in the case of third moments). Both effects impact mostly small scales. In this work, source clustering in our simulations was forward modelled following the prescription presented in [42] (see Eq. 2). This implementation assumes a linear galaxy-matter bias for our sample. Furthermore, for simplicity, we also chose not to marginalise over such a bias, instead fixing its value to unity. We took some precautions to minimise the effect of source clustering in case our source clustering model does not faithfully reproduce the effects on data (which might happen, for instance, if the galaxy-matter bias of the source was different from unity). In [42], the authors pointed out that for third moments the largest effect due to source clustering is related to the spurious signal-noise correlations, and that this can be removed completely by subtracting from the third moments estimators specific moments involving combinations of the observed noisy maps and noise-only maps (see §III A). For the other statistics used in this work, we tested that source clustering effects are most noticeable for WPH $S01$ and WPH $C01$, and negligible for the other statistics. For WPH $S01$ and WPH $C01$, therefore, we adopted a noise-subtraction procedure similar to the one applied to third moments (see §III B, III C), which we empirically found to reduce the impact of source clustering on the measurements.

In order to test the impact of any potential mismodelling of source clustering effects on our results, we analysed two sets of maps generated assuming a galaxy-matter bias $b = 0.5$ or $b = 1.5$ instead of unity. We verified that in none of our combination of summary statistics did the bias in the S_8-Ω_m plane exceed 0.10σ. This means that the impact on cosmological parameters is safely negligible and that our modelling of source clustering is sufficiently accurate that small scales need not be removed from our analysis.

C. Impact of additive biases due to PSF errors

We assess here the degree of contamination in our data vector resulting from the inclusion of additive biases associated with the misestimation of the Point Spread Function (PSF). The misestimation of the PSF can introduce additional biases in the measured shapes of galaxies, leading to deviations from their true values:

$$\boldsymbol{\gamma}^{\mathbf{est}} = \boldsymbol{\gamma} + \delta\boldsymbol{e}^{\mathrm{sys}}_{\mathrm{PSF}} + \delta\boldsymbol{e}^{\mathrm{noise}}. \tag{24}$$

To quantify these unwanted contributions, we can employ a model that accounts for the errors in PSF modelling and use a catalogue of 'reserved' stars. These reserved stars are not used in training the PSF model and serve as a reference to characterize the spurious effects accurately. We follow [55] and [43] by assuming that

$$\delta\boldsymbol{e}^{\mathrm{sys}}_{\mathrm{PSF}} = \alpha\boldsymbol{e}_{\mathrm{model}} + \beta\,(\boldsymbol{e}_* - \boldsymbol{e}_{\mathrm{model}}) + \eta\left(\boldsymbol{e}_*\frac{T_* - T_{\mathrm{model}}}{T_*}\right), \tag{25}$$

where α, β, and η are coefficients estimated from data, $\boldsymbol{e}_*$ is the PSF ellipticity measured directly using the reserved stars catalogue, T_{model} is the modelled PSF size, and T_* is the PSF size measured from the reserved stars catalogue. The coefficients α, β, and η for the DES Y3 shape catalogue for the four tomographic bins are provided in [7].

We use an empirical method to estimate the contribution of PSF additive biases to the summary statistics used in this work. We first created maps of $\boldsymbol{e}_{\mathrm{model}}$, $\boldsymbol{e}_*$, and $\boldsymbol{e}_*\frac{T_*-T_{\mathrm{model}}}{T_*}$ from the reserved stars catalogue. Using the estimated values for α, β, and η, we then created maps of $\delta\boldsymbol{e}^{\mathrm{sys}}_{\mathrm{model}}$, one for each tomographic bin. We added these systematic maps to a set of simulated maps at the fiducial cosmology, and proceeded to compute the summary statistics and to analyse the measurement with our LFI pipeline. We repeated the same procedure on maps with no PSF additive biases, and compared the two analyses at the level of the constraints in the S_8-Ω_m plane. We

verified that in none of our combinations of summary statistics did the bias in the S_8-Ω_m plane exceed 0.10σ, indicating that PSF modelling errors are negligible for the range of scales used in this work.

VI. END-TO-END TESTS ON SIMULATIONS

Having verified that all the scales used in our analysis are safe against a number of systematics, we next verify that we are able to recover the true cosmology of a set of simulations that have not been used to build our pipeline. To this end, we use 400 independent DES Y3 mock catalogues produced with the `CosmoGridV1` simulations. Each mock has the same cosmology; we further assume no intrinsic alignment, while for the other nuisance parameters (shear calibration and redshift uncertainties) we assume values at the centre of the priors. We measure all the summary statistics in the mocks, and then we average them, to reduce the impact of noise.

Our LFI analysis marginalises over seven cosmological parameters, assuming a νwCDM model; moreover, it marginalises over multiplicative shear bias (four parameters), intrinsic alignment (two parameters), and redshift distributions, as summarised in Table I. In addition to these parameters, we will also quote results in terms of the S_8 parameter, defined as

$$S_8 \equiv \sigma_8(\Omega_m/0.3)^\alpha \,. \tag{26}$$

The value of α can be chosen so that S_8 best constrains the degeneracy between Ω_m and σ_8. However, the summary statistics considered in this work have different directions and so there is no value of α that simultaneously optimises all. For sake of simplicity we adopt $\alpha = 0.5$. We also quote a Figure-of-Merit (FoM), defined for S_8, Ω_m, and their covariance:

$$\mathrm{FoM}_{S_8,\Omega_m} = \left(\det(C_{S_8,\Omega_m})\right)^{-0.5} \tag{27}$$

Fig. 11 shows the posteriors for S_8, Ω_m, and σ_8 for a combination of different summary statistics; posteriors for other summary statistics are shown in Fig. 12 for S_8 and Ω_m. In Fig. 11, 'All' means that all the summary statistics are combined, except for WPHG, as we found it does not add additional information compared to second moments alone. For this reason we also chose to always use second moments as a default Gaussian statistic when combining with other non-Gaussian probes. Individual parameter constraints, together with the $\mathrm{FoM}_{S_8,\Omega_m}$, are reported in Table V.

From Figs. 11 and 12 it can be noted that non-Gaussian statistics such as third moments and WPH S01 and C01 are characterised by a slightly different degeneracy tilt in the σ_8-Ω_m plane compared to second moments. This distinction also becomes apparent in the S_8-Ω_m plane, as the posteriors deviate from alignment with the S_8 axis. For other non-Gaussian statistics, such as ST1, ST2, or WPH S00, this is less evident, and is probably due to their being highly correlated with the second moments.

When all the summary statistics are combined, the gain in terms of constraining power over the standard Gaussian statistics (either second moments or WPHG) is substantial: the constraints on S_8 improve by ~ 25 per cent, whereas the gain in terms of $\mathrm{FoM}_{S_8,\Omega_m}$ is ~ 90 per cent, i.e. almost double. This level of improvement is expected, and is due to the additional non-Gaussian information probed by the non-Gaussian WPH moments, ST, and third moments, and the degeneracy breaking.

When looking at the individual probes, we find that the WPHG are slightly less constraining than second moments alone (~ 10 per cent less constraining on the FoM). As they both probe the power spectrum of the maps, this indicates that the spacing between the wavelet filters used for the WPHG (where each filter scale is double the size of the one preceding it) is inferior to the spacing of the top hat filters used for the second moments filters (where we considered more intermediate scales). A similar results was also found by [111] using simulations. This problem could be mitigated by also introducing additional scales for the wavelet filters; we leave this exploration to future works. We also find that ST1 and ST2, either individually or combined, are not as constraining as second moments (~ 10 per cent less constraining on the FoM when combined), despite appearing to be highly correlated (Fig. 5), and despite being characterised by a high signal-to-noise (Table II). We found that this is due to a non-optimal information extraction from cross-bins (Eq. 21); including the cross-maps in the data vector for ST improves the constraints only a small amount, whereas second moments or WPHG significantly improve their constraints when cross-bins are included in the data vector. This would suggest a need to explore alternative ways of incorporating cross-correlation information among diverse fields within the ST framework. Alternatively, this lends support to the idea of employing WPH, which naturally facilitates the correlation of distinct fields.

Next, of the three categories of non-Gaussian statistics examined in this study, the strongest performance – in terms of constraining power when combined with second moments – is exhibited by WPH, with ST following, and third moments trailing. Nevertheless, the combination of all the different statistics continues to enhance the constraints, underscoring that each statistic delves into slightly distinct information.

We next look into the constraints for the other parameters varied in this analysis.

Fig. 13 shows the constraints on S_8 and A_{IA} (the amplitude of IA) for some of the summary statistics (and their combinations) considered here (see also Table V). The amplitude of IA is one of the other main parameters constrained by weak lensing probes [25]. The posteriors recover the correct value ($A_{IA} = 0$); interestingly, whenever second moments are combined with any of the non-Gaussian statistic considered here, constraints on A_{IA} are improved, up to almost 20 per cent. The parameter η_{IA} (which controls the redshift evolution of the IA amplitude) is not very well constrained as, for $A_{IA} = 0$, any

Table V. Constraints on various parameters for different summary statistics and their combinations. All the other parameters are marginalised over (see Table I for a list of parameters and their priors). For each parameter we report the 68 per cent confidence interval; numbers in parentheses refer to the percentage gain (or loss) with respect to the constraints from the second moments. Note that the improvement on the FOM in the last column is the most meaningful metric of a method's statistical power.

Summary Statistic(s)	$\sigma(S_8)$ [x100]	$\sigma(\sigma_8)$ [x100]	$\sigma(\Omega_m)$ [x100]	$\sigma(w)$ [x10]	$\sigma(A_{IA})$ [x10]	FoM(S_8, Ω_m) -
2nd moments	2.7	5.3	3.4	1.3	4.4	904
2nd + 3rd moments	2.6(+ 3%)	5.0(+ 6%)	3.4(-0%)	1.3(-3%)	4.2(+ 5%)	1035(+15%)
2nd moments + ST	2.7(+ 2%)	4.3(+19%)	3.0(+12%)	1.2(+11%)	4.4(+ 0%)	1245(+38%)
2nd moments + WPH	2.4(+11%)	4.4(+18%)	2.9(+15%)	1.1(+15%)	3.9(+10%)	1385(+53%)
2nd moments + ST + WPH	2.0(+25%)	4.4(+18%)	2.9(+15%)	1.2(+ 9%)	3.6(+17%)	1684(+86%)
2nd + 3rd moments + ST + WPH	2.0(+25%)	3.9(+26%)	2.9(+14%)	1.2(+12%)	3.6(+17%)	1733(+92%)

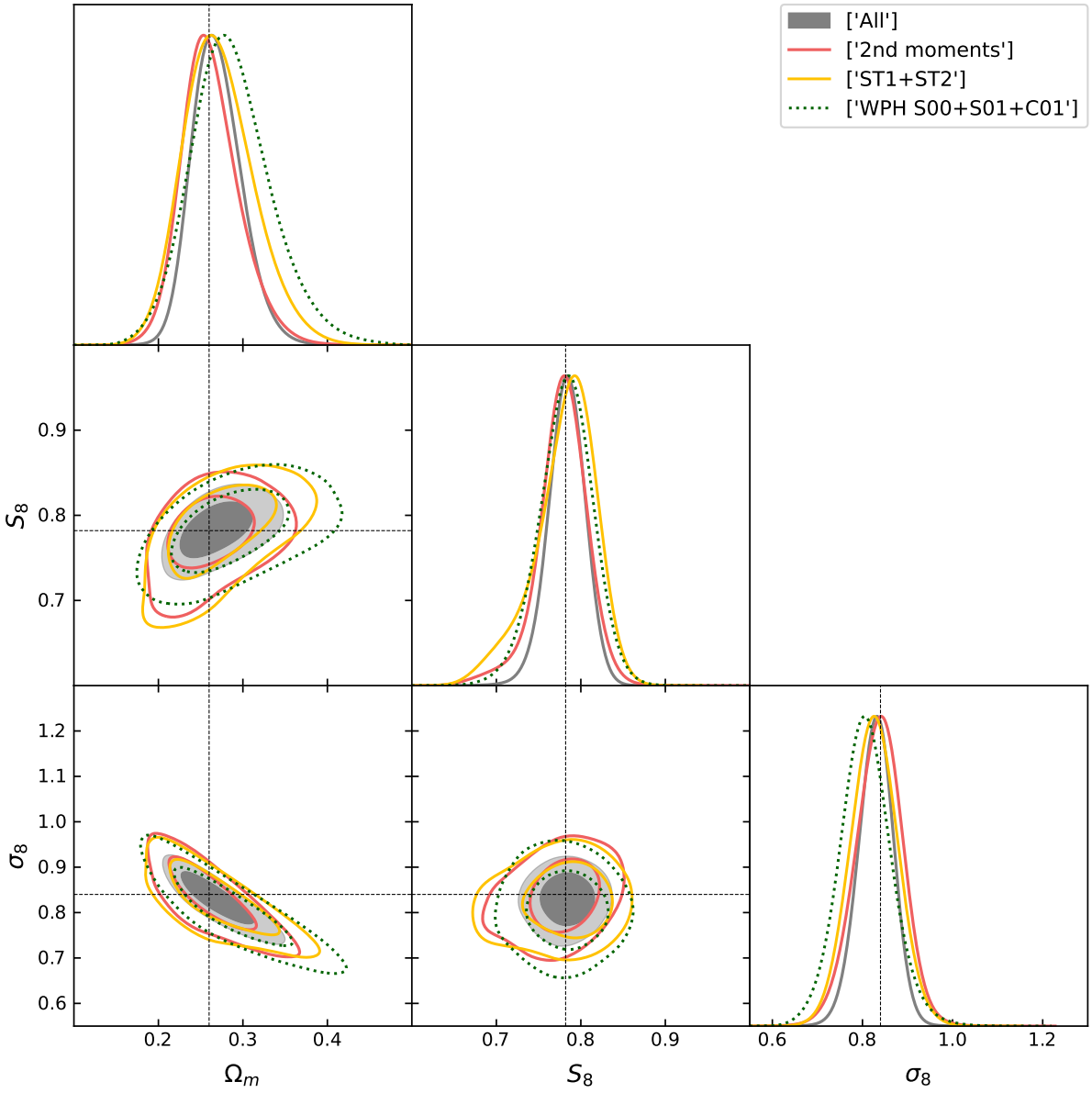

Figure 11. Posterior distributions of the cosmological parameters Ω_m, S_8, and σ_8, for different summary statistics and their combinations, as measured in `CosmoGridV1` simulations. The dotted black lines indicate the values of the cosmological parameters in the simulations. The two-dimensional marginalised contours in these figures show the 68 per cent and 95 per cent confidence levels.

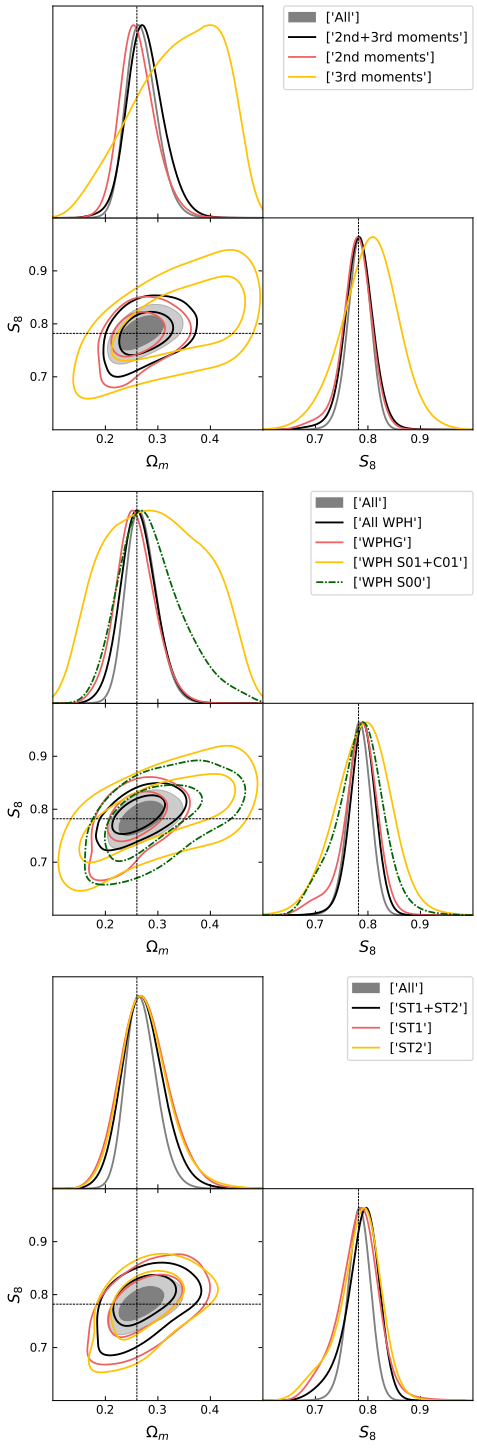

Figure 12. Posterior distributions of the cosmological parameters Ω_m and S_8, for different summary statistics and their combinations, as measured in `CosmoGridV1` simulations. The dotted black lines indicate the values of the cosmological parameters in the simulations. The two-dimensional marginalised contours in these figures show the 68 per cent and 95 per cent confidence levels.

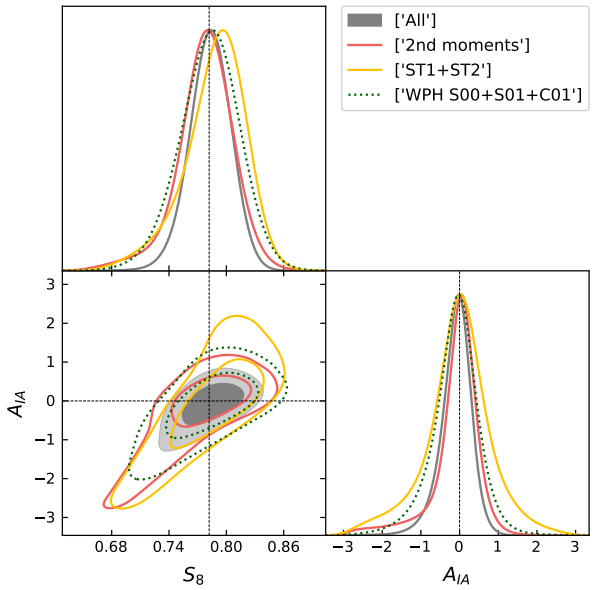

Figure 13. Posterior distributions of the cosmological parameters S_8 and A_IA, for different summary statistics and their combinations, as measured in `CosmoGridV1` simulations. The dotted black lines indicate the values of the cosmological parameters in the simulations. The two-dimensional marginalised contours in these figures show the 68 percent and 95 percent confidence levels. Note that there is significant improvement from using non-Gaussian statistics in the 95 percent confidence levels but less so in the 68 percent levels.

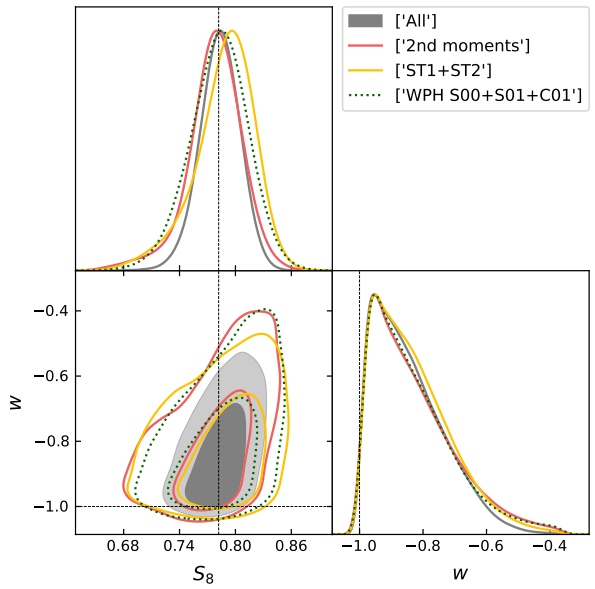

Figure 14. Posterior distributions of the cosmological parameters S_8 and w, for different summary statistics and their combinations, as measured in `CosmoGridV1` simulations. The dotted black lines indicate the values of the cosmological parameters in the simulations. The two-dimensional marginalised contours in these figures show the 68 per cent and 95 per cent confidence levels.

value of $\eta_{\rm IA}$ would provide an equally good fit. Recall that we used a slightly simpler IA model than the fiducial DES Y3 analysis: ours does not include tidal-torque terms (because our current pipeline lacks the capability to compute these terms). It is possible that the enhancement in constraining power resulting from the incorporation of the non-Gaussian statistics of these extra IA terms might be different the ones obtained for $A_{\rm IA}$ and $\eta_{\rm IA}$; we defer this investigation to future work.

Fig. 14 shows the constraints on S_8 and w (see also Table V). The `CosmoGridV1` simulations used here have been produced assuming a ΛCDM cosmology ($w = -1$): correctly, Fig. 14 shows the posteriors skewed towards the edge of the prior. Despite these posteriors being partially prior-dominated, the combination of different non-Gaussian statistics improves the constraints on w with respect to second moments by roughly 10 per cent. We also analysed the posterior distributions of the four parameters describing the shear multiplicative biases (m_i), and four parameters (Δz_i) describing the shift in the mean redshift of the $n(z)$. The Δz_i have been estimated for each of the multiple $n(z)$ realisations that have been used to produce our simulated maps with respect to the fiducial DES Y3 $n(z)$ given by the mean of all these realisations. As for the priors on Δz_i, we assumed them be Gaussian with zero mean and standard deviations equal to the spread of the shifts. These parameters are usually dominated by their priors, and, typically, conventional Gaussian statistics struggle to improve over these prior constraints. Some recent studies have pointed out the potential of non-Gaussian statistics for self-calibration, as evidenced by their ability to enhance precision beyond prior limitations in such parameters [88]. Indeed, we already saw this effect for $A_{\rm IA}$. The posteriors for Δz_i and m_i, however, were basically the same as their priors; we noted only a small improvement for the Δz_i corresponding to the second, third, and fourth bins by 5-10 per cent for the combination of all the summary statistics.

VII. CONCLUSIONS

In this methodology paper, we have presented an end-to-end simulation-based cosmological analysis of a set of Gaussian and non-Gaussian weak lensing statistics using detailed mock catalogues of the first three years of data of the Dark Energy Survey. Our main goals are to show the constraining power of wavelet based non-Gaussian statistics and to validate a simulation based inference framework for a broad class of statistics for lensing surveys.

We considered the following summary statistics of weak lensing mass maps: 1) second and third moments; 2) wavelet phase harmonics (WPH); 3) the scattering transform (ST). Second moments are Gaussian statistics, whereas third moments probe additional non-Gaussian information of the fields. The WPH moments are second moments of smoothed weak lensing mass maps that have undergone a non-linear transformation, allowing for the exploration of the non-Gaussian features of the field. The ST coefficients are built through a series of smoothing and modulus operations applied to the input field, followed by an average. The WPH and ST are often linked to convolutional neural networks (CNNs) because the definition of the statistics bears similarities to the architecture of CNNs (but note the latter requires training data). They capture both Gaussian and non-Gaussian features of the fields; however, being only first or second order in the input data, they are generally more robust to noise than higher order moments. Moreover, in our implementation of the WPH and ST, we considered maps smoothed by directional wavelets, whereas for moments we only considered isotropic top-hat filters.

Our analysis is fully based on simulations. We produced 791 full-sky N-body simulations, spanning seven cosmological parameters assuming a νwCDM cosmology: $\Omega_{\rm m}$, σ_8, n_s, h_{100}, $\Omega_{\rm b}$, w, m_ν. Using the N-body full-sky simulations, we generated almost 13000 pseudo-independent DES Y3 weak lensing mock mass maps, which we used for our inference pipeline. Our mock mass maps implement realistic masks, noise variations, source clustering of the sources, and include the following astrophysical observational systematic effects: intrinsic alignments, shear calibration, and redshift calibration biases. Our analysis is tomographic, i.e. we forward model the four tomographic bins and maps into which the DES Y3 weak lensing sample is divided.

We implemented a neural network compression of the summary statistics, and we estimated the parameter posteriors using a likelihood-free-inference (LFI) approach, with a combination of Gaussian Mixture Density Networks and Masked Autoregressive Flows to estimate the likelihood surface from our mocks. We extensively validated our pipeline, testing the size of the posteriors with a coverage probability test, and comparing the posterior obtained from the LFI pipeline against a theory-based and Gaussian likelihood approach for the special case of Gaussian statistics (i.e. the power spectrum of the maps).

We tested that the scales used in this work were not affected by systematics not properly modelled in our simulations: namely, baryonic feedback effects, PSF modelling errors, and differences in the prescriptions used to model source clustering. Finally, we tested our pipeline on a set of independent simulations that have not been used in our training process, demonstrating we could recover the true values of the cosmological parameters of the simulation.

Of the three combinations of 'non-Gaussian statistic plus second moment' examined, WPH exhibits the strongest constraining power, followed by ST, and then third moments. The combination of all the different statistics continues to enhance the constraints, underscoring that each statistic delves into slightly distinct information. In particular, we found that when all the summary statistics are combined, the constraints on S_8, $\Omega_{\rm m}$, and on the Figure-Of-Merit $\mathrm{FoM}_{S_8,\Omega_{\rm m}}$ are improved by roughly 25 per cent, 15 per cent, and 90 percent, re-

spectively, over the constraints from second moments. Similar gains are found on w (∼15 percent), and on the amplitude of intrinsic alignment (∼20 percent).

This work highlights the importance of analysing probes of higher order statistics to improve the cosmological constraints, and showcases the power of a full simulation-based framework to efficiently model and combine different non-Gaussian probes. Here we targeted the analysis at the third year (Y3) data from the Dark Energy Survey (DES), but the methodological advances presented here are suitable for application to Stage IV surveys from Euclid, Rubin-LSST, and Roman, once any necessary additional validation is caried out on mock catalogues for each survey. In a companion paper (Gatti et al., in prep.) we present an application to the DES Year 3 data.

ACKNOWLEDGEMENTS

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

Based in part on observations at Cerro Tololo Inter-American Observatory at NSF's NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under grants ESP2017-89838, PGC2018-094773, PGC2018-102021, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

[1] Abbott, T. M. C., Aguena, M., Alarcon, A., et al., 2022, Phys. Rev. D, 105, 2, 023520, arXiv:2105.13549
[2] Aghanim, N., Akrami, Y., Ashdown, M., et al., 2020, Astronomy & Astrophysics, 641, A6
[3] Aihara, H., Arimoto, N., Armstrong, R., et al., 2018, PASJ, 70, S4
[4] Ajani, V., Peel, A., Pettorino, V., Starck, J.-L., Li, Z., Liu, J., 2020, Phys. Rev. D, 102, 10, 103531
[5] Allys, E., Marchand, T., Cardoso, J. F., Villaescusa-Navarro, F., Ho, S., Mallat, S., 2020, Phys. Rev. D, 102, 10, 103506
[6] Alsing, J., Wandelt, B., Feeney, S., 2018, MNRAS, 477, 2874
[7] Amon, A., Gruen, D., Troxel, M. A., et al., 2022, Phys. Rev. D, 105, 2, 023514
[8] Anbajagane, D., Chang, C., Banerjee, A., et al., 2023, arXiv e-prints, arXiv:2308.03863
[9] Aricò, G., Angulo, R. E., Hernández-Monteagudo, C., et al., 2020, MNRAS, 495, 4, 4800
[10] Asgari, M., Lin, C.-A., Joachimi, B., et al., 2021, A&A, 645, A104
[11] Banerjee, A., Abel, T., 2023, MNRAS, 519, 4, 4856
[12] Barthelemy, A., Codis, S., Uhlemann, C., Bernardeau, F., Gavazzi, R., 2020, MNRAS, 492, 3, 3420
[13] Bishop, C. M., 1994, Mixture density networks, Tech. rep., Aston University
[14] Blazek, J. A., MacCrann, N., Troxel, M. A., Fang, X., 2019, Phys. Rev. D, 100, 10, 103506

[15] Boruah, S. S., Lavaux, G., Hudson, M. J., 2022, MNRAS, 517, 3, 4529
[16] Boyle, A., Uhlemann, C., Friedrich, O., et al., 2021, MNRAS, 505, 2, 2886
[17] Bridle, S., King, L., 2007, New Journal of Physics, 9, 444
[18] Bruna, J., Mallat, S., 2013, arXiv e-prints, arXiv:1311.0407
[19] Chang, C., Pujol, A., Mawdsley, B., et al., 2018, MNRAS, 475, 3165
[20] Chang, C., Vikram, V., Jain, B., et al., 2015, Phys. Rev. Lett., 115, 5, 051301
[21] Cheng, S., Ting, Y.-S., Ménard, B., Bruna, J., 2020, MNRAS, 499, 4, 5902
[22] Cohen, A., Ryan, R. D., 1995, Wavelets and multiscale signal processing, Chapman & Hall
[23] Collaboration, D. E. S., 2016, MNRAS, 460, 2, 1270
[24] Cordero, J. P., Harrison, I., Rollins, R. P., et al., 2022, MNRAS, 511, 2, 2170
[25] Dacunha, T., Raveri, M., Park, M., Doux, C., Jain, B., 2022, Phys. Rev. D, 105, 6, 063529
[26] Dalal, R., Li, X., Nicola, A., et al., 2023, arXiv e-prints, arXiv:2304.00701
[27] Dietrich, J. P., Hartlap, J., 2010, MNRAS, 402, 2, 1049
[28] Doux, C., Jain, B., Zeurcher, D., et al., 2022, MNRAS, 515, 2, 1942
[29] Einstein, A., 1936, Science, 84, 2188, 506
[30] Euclid Collaboration, Ajani, V., Baldi, M., et al., 2023, A&A, 675, A120, arXiv:2301.12890
[31] Euclid Collaboration, Knabenhans, M., Stadel, J., et al., 2021, MNRAS, 505, 2, 2840
[32] Feldbrugge, J., van Engelen, M., van de Weygaert, R., Pranav, P., Vegter, G., 2019, JCAP, 2019, 9, 052
[33] Fluri, J., Kacprzak, T., Lucchi, A., Schneider, A., Refregier, A., Hofmann, T., 2022, Phys. Rev. D, 105, 8, 083518
[34] Fluri, J., Kacprzak, T., Lucchi, A., et al., 2019, Phys. Rev. D, 100, 6, 063514
[35] Fluri, J., Kacprzak, T., Refregier, A., Amara, A., Lucchi, A., Hofmann, T., 2018, Phys. Rev. D, 98, 12, 123518
[36] Foreman, S., Coulton, W., Villaescusa-Navarro, F., Barreira, A., 2020, MNRAS, 498, 2, 2887
[37] Foreman-Mackey, D., Hogg, D. W., Lang, D., Goodman, J., 2013, PASP, 125, 306
[38] Fosalba, P., et al., 2015, MNRAS, 447, 2, 1319
[39] Gatti, M., Chang, C., Friedrich, O., et al., 2020, MNRAS, 498, 3, 4060
[40] Gatti, M., Giannini, G., Bernstein, G. M., et al., 2022, MNRAS, 510, 1, 1223
[41] Gatti, M., Jain, B., Chang, C., et al., 2022, Phys. Rev. D, 106, 8, 083509
[42] Gatti, M., Jeffrey, N., Whiteway, L., et al., 2023, arXiv e-prints, arXiv:2307.13860
[43] Gatti, M., Sheldon, E., Amon, A., et al., 2021, MNRAS, 504, 3, 4312
[44] Germain, M., Gregor, K., Murray, I., Larochelle, H., 2015, in International Conference on Machine Learning, 881–889
[45] Górski, K. M., Hivon, E., Banday, A. J., et al., 2005, ApJ, 622, 759
[46] Grewal, N., Zuntz, J., Tröster, T., Amon, A., 2022, The Open Journal of Astrophysics, 5, 1, 13, arXiv:2206.03877
[47] Handley, W. J., Hobson, M. P., Lasenby, A. N., 2015, MNRAS, 450, L61
[48] Handley, W. J., Hobson, M. P., Lasenby, A. N., 2015, MNRAS, 453, 4, 4384
[49] Harnois-Déraps, J., Martinet, N., Reischke, R., 2022, MNRAS, 509, 3, 3868
[50] Heavens, A. F., Jimenez, R., Lahav, O., 2000, MNRAS, 317, 965
[51] Heavens, A. F., Sellentin, E., de Mijolla, D., Vianello, A., 2017, MNRAS, 472, 4244
[52] Heydenreich, S., Brück, B., Burger, P., et al., 2022, A&A, 667, A125
[53] Heydenreich, S., Brück, B., Harnois-Déraps, J., 2021, A&A, 648, A74
[54] Huff, E., Mandelbaum, R., 2017, arXiv e-prints, 1702.02600
[55] Jarvis, M., Sheldon, E., Zuntz, J., et al., 2016, MNRAS, 460, 2245
[56] Jeffrey, N., Alsing, J., Lanusse, F., 2021, MNRAS, 501, 1, 954
[57] Jeffrey, N., Gatti, M., Chang, C., et al., 2021, MNRAS, 505, 3, 4626
[58] Jeffrey, N., Wandelt, B. D., 2020, Third Workshop on Machine Learning and the Physical Sciences, NeurIPS 2020, arXiv:2011.05991
[59] Kacprzak, T., Fluri, J., Schneider, A., Refregier, A., Stadel, J., 2023, JCAP, 2023, 2, 050
[60] Kacprzak, T., Kirk, D., Friedrich, O., et al., 2016, MNRAS, 463, 3653
[61] Kaiser, N., Squires, G., 1993, ApJ, 404, 441
[62] Kratochvil, J. M., Haiman, Z., May, M., 2010, Phys. Rev. D, 81, 4, 043519
[63] Kratochvil, J. M., Lim, E. A., Wang, S., Haiman, Z., May, M., Huffenberger, K., 2012, Phys. Rev. D, 85, 10, 103513
[64] Krause, E., Fang, X., Pandey, S., et al., 2021, arXiv e-prints, arXiv:2105.13548
[65] Kuijken, K., Heymans, C., Hildebrandt, H., et al., 2015, MNRAS, 454, 4, 3500
[66] Laureijs, R., Amiaux, J., Arduini, S., et al., 2011, arXiv e-prints, arXiv:1110.3193
[67] Li, X., Zhang, T., Sugiyama, S., et al., 2023, arXiv e-prints, arXiv:2304.00702
[68] Liu, J., Petri, A., Haiman, Z., Hui, L., Kratochvil, J. M., May, M., 2015, Phys. Rev. D, 91, 6, 063507
[69] LSST Science Collaboration, Abell, P. A., Allison, J., et al., 2009, arXiv e-prints, arXiv:0912.0201
[70] Lu, T., Haiman, Z., Li, X., 2023, MNRAS, 521, 2, 2050
[71] MacCrann, N., Becker, M. R., McCullough, J., et al., 2022, MNRAS, 509, 3, 3371, arXiv:2012.08567
[72] Madhavacheril, M. S., Qu, F. J., Sherwin, B. D., et al., 2023, arXiv e-prints, arXiv:2304.05203
[73] Mallat, S., 1999, A wavelet tour of signal processing, Elsevier
[74] Mallat, S., 2011, arXiv e-prints, arXiv:1101.2286
[75] Mallat, S., 2016, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374, 2065, 20150203
[76] Mallat, S., et al., 2020, Information and Inference: A Journal of the IMA, 9, 3, 721, ISSN 2049-8772
[77] Martinet, N., Schneider, P., Hildebrandt, H., et al., 2018, MNRAS, 474, 1, 712
[78] Myles, J., Alarcon, A., Amon, A., et al., 2021, MNRAS, 505, 3, 4249
[79] Oguri, M., Miyazaki, S., Hikage, C., et al., 2018, PASJ, 70, S26

[80] Papamakarios, G., Pavlakou, T., Murray, I., 2017, Advances in neural information processing systems, 30
[81] Parroni, C., Cardone, V. F., Maoli, R., Scaramella, R., 2020, A&A, 633, A71
[82] Parroni, C., Tollet, É., Cardone, V. F., Maoli, R., Scaramella, R., 2021, A&A, 645, A123
[83] Peel, A., Pettorino, V., Giocoli, C., Starck, J.-L., Baldi, M., 2018, A&A, 619, A38
[84] Petri, A., Liu, J., Haiman, Z., May, M., Hui, L., Kratochvil, J. M., 2015, Phys. Rev. D, 91, 10, 103511
[85] Porqueres, N., Heavens, A., Mortlock, D., Lavaux, G., 2022, MNRAS, 509, 3, 3194
[86] Porth, L., Smith, R. E., 2021, MNRAS, 508, 3, 3474
[87] Potter, D., Stadel, J., Teyssier, R., 2017, Computational Astrophysics and Cosmology, 4, 1, 2
[88] Pyne, S., Joachimi, B., 2021, MNRAS, 503, 2, 2300
[89] Ribli, D., Pataki, B. Á., Csabai, I., 2019, Nature Astronomy, 3, 93
[90] Riess, A. G., Yuan, W., Macri, L. M., et al., 2022, The Astrophysical Journal Letters, 934, 1, L7
[91] Schaye, J., Dalla Vecchia, C., Booth, C. M., et al., 2010, MNRAS, 402, 1536
[92] Schmidt, F., Rozo, E., Dodelson, S., Hui, L., Sheldon, E., 2009, ApJ, 702, 1, 593
[93] Schneider, A., Teyssier, R., 2015, JCAP, 2015, 12, 049
[94] Schneider, A., Teyssier, R., Stadel, J., et al., 2019, JCAP, 2019, 3, 020
[95] Schneider, P., van Waerbeke, L., Mellier, Y., 2002, A&A, 389, 729
[96] Secco, L. F., Samuroff, S., Krause, E., et al., 2022, Phys. Rev. D, 105, 2, 023515
[97] Shan, H., Liu, X., Hildebrandt, H., et al., 2018, MNRAS, 474, 1, 1116
[98] Sheldon, E. S., Huff, E. M., 2017, ApJ, 841, 24
[99] Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., Oguri, M., 2012, ApJ, 761, 152
[100] Thiele, L., Hill, J. C., Smith, K. M., 2020, Phys. Rev. D, 102, 12, 123545
[101] Valageas, P., 2014, A&A, 561, A53
[102] Valogiannis, G., Dvorkin, C., 2022, Phys. Rev. D, 106, 10, 103509
[103] Valogiannis, G., Dvorkin, C., 2022, Phys. Rev. D, 105, 10, 103534
[104] van Daalen, M. P., Schaye, J., Booth, C. M., Dalla Vecchia, C., 2011, MNRAS, 415, 3649
[105] Van Den Berg, J., 1999, Wavelets in Physics, Cambridge University Press
[106] Van Waerbeke, L., Benjamin, J., Erben, T., et al., 2013, MNRAS, 433, 3373
[107] Vicinanza, M., Cardone, V. F., Maoli, R., Scaramella, R., Er, X., 2016, arXiv e-prints, arXiv:1606.03892
[108] Vicinanza, M., Cardone, V. F., Maoli, R., Scaramella, R., Er, X., 2018, Phys. Rev. D, 97, 2, 023519
[109] Vicinanza, M., Cardone, V. F., Maoli, R., Scaramella, R., Er, X., Tereno, I., 2019, Phys. Rev. D, 99, 4, 043534
[110] Vikram, V., Chang, C., Jain, B., et al., 2015, Phys. Rev. D, 92, 2, 022006
[111] Zürcher, D., Fluri, J., Ajani, V., Fischbacher, S., Refregier, A., Kacprzak, T., 2023, MNRAS, 525, 1, 761, arXiv:2206.01450
[112] Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., Refregier, A., 2021, JCAP, 2021, 1, 028
[113] Zürcher, D., Fluri, J., Sgier, R., et al., 2022, MNRAS, 511, 2, 2075

Appendix A: Noise properties of the simulations

In this Appendix, we conduct several sanity checks to evaluate the noise characteristics of our simulations. Specifically, we ensure that the noise properties of the simulations encompass those of the actual data. The noise properties of the simulations should be mildly cosmology dependent, due to source clustering effects (see § II B 3).

To perform this test, we consider the following statistics: moments (second, third, and fourth order) and cumulative distribution functions (CDFs). The CDFs [8, 11] for a given field are defined as the fraction of circles that have an enclosed value of the field larger than a given threshold:

$$\mathrm{CDF}\,(\theta, k = P(\kappa_\theta > k)), \tag{A1}$$

where k is the threshold. The CDFs can be formally shown to contain all volume integrals of higher-order functions [11]. We measure the CDFs across ten smoothing scales, spaced logarithmically between 3.2 and 200 arcmin; for each scale, we use five thresholds $k \in [-20, -6, -2, 0, 2] \times 10^{-3}$.

In Fig. 15 we compare a) moments and CDFs from data to b) moments and CDFs from simulations; there is a good match, indicating that the noise properties of our simulations reproduce well the noise of the data.

Appendix B: Neural compression vs. MOPED compression

In this work we opted for a neural compression scheme to compress our summary statistics. Other compression methods exist; the most notable is the MOPED algorithm [50], which is lossless when the likelihood is Gaussian and the covariance matrix of the observables has a negligible dependence on the parameters. The neural network implemented in this work is in principle more powerful and general than the MOPED compression, as it does not make any assumption about the Gaussianity of the likelihood, nor about any dependence on the model parameters. Even if desired, we would not have been able to implement the MOPED compression for most of the statistics, as doing so would have required an estimate of the derivative of the model with respect to the parameters; such derivatives are available in closed form for analytical models, or via finite difference for observables where the model is estimated from simulations – the Gower St simulations, however, do not allow us to estimate derivatives through finite differences.

In this Appendix we compare the neural network compression with the MOPED compression for the only summary statistic for which we have an analytical model, i.e. the second moments. We also know that for second moments the likelihood should be fairly Gaussian, and the covariance should only weakly depend on parameters, so the MOPED compression should be close to lossless. We therefore compute the derivatives needed for the MOPED compression using the analytic model from

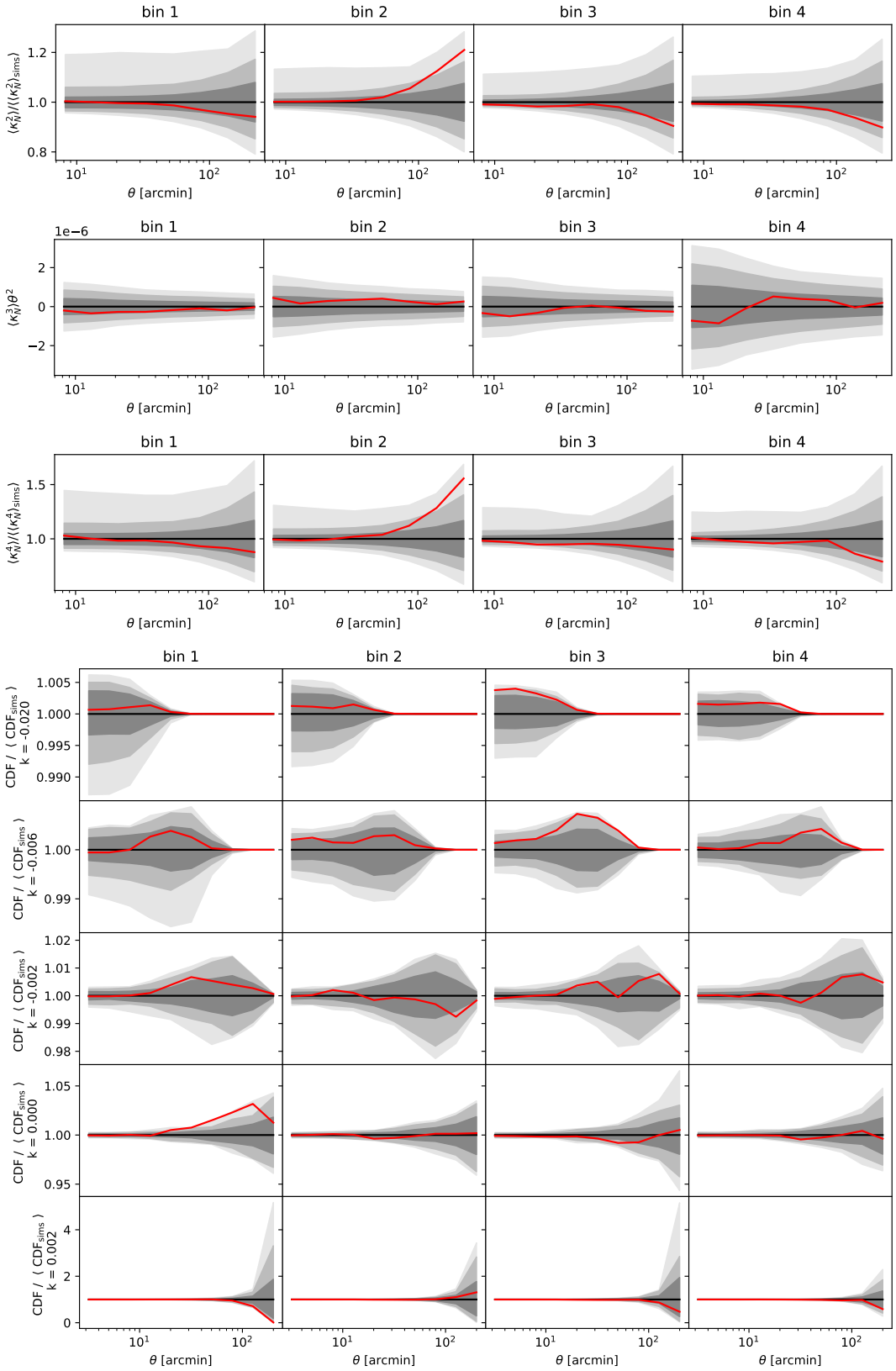

Figure 15. Moments and CDFs of noise-only maps (grey shaded regions) in the Gower St simulations compared to the same quantities in data (red lines). The three different grey shaded regions encompass the 68, 95, and 99.5 percentiles spanned by the noise moments and CDFs in the simulations.

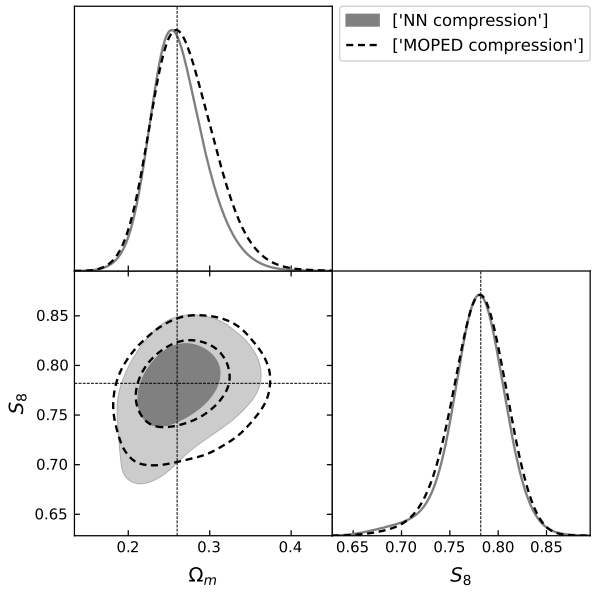

Figure 16. Posterior distribution of the cosmological parameters Ω_m and S_8 as measured in `CosmoGridV1` simulations. The two different posteriors have been obtained by analysing the second moments data vector compressed both with our fiducial neural network compression and with the alternative MOPED compression.

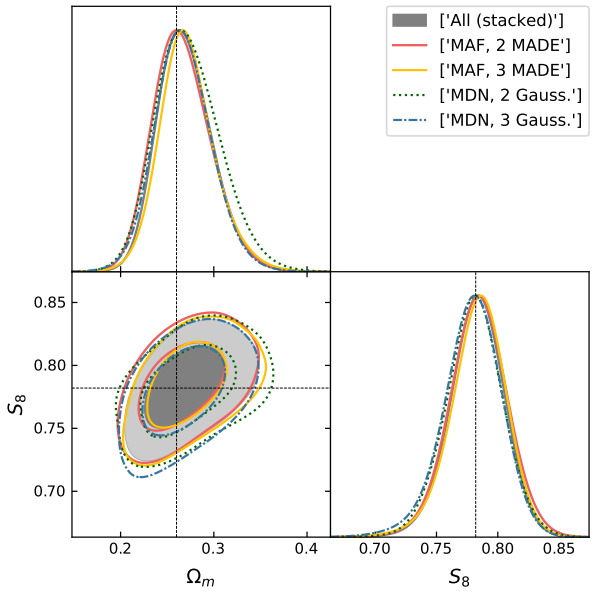

Figure 17. Posterior distributions of the cosmological parameters Ω_m and σ_8 for the combination of all the summary statistics considered in this work, as measured in `CosmoGridV1` simulations. We show the different posteriors as estimated by the different NDEs used in this work; we also show their stacked combination (the fiducial setup used in the other Figures of this paper). The dotted black lines indicate the values of the cosmological parameters in the simulations. The two-dimensional marginalised contours in these figures show the 68 per cent and 95 per cent confidence levels.

[39]; for the covariance, we estimate it from the 400 measurements of the second moments in the `CosmoGridV1` simulations. Fig. 16 shows the posteriors obtained using our pipeline, compressing second moments either with the neural network or with the MOPED compression. Results are fairly similar, with the neural network compression delivering only slightly tighter contours.

Appendix C: NDEs and parameters posterior

In this work we used four different neural density estimators (NDEs) to estimate the posteriors. In particular, we used two different Gaussian Mixture Density Networks (MDNs) and two different Masked Autoencoders for Distribution Estimation (MADEs). Whenever we showed a posterior or reported the constraints on some parameters in this work, we always obtained these by stacking the four different NDEs. Assuming all the NDEs are flexible enough to describe our likelihood surface, they should all agree in the limit in which the number of simulations used for training becomes large. Fig. 17 shows the posteriors obtained by each individual NDE for our most constraining case (i.e. the combination of all summary statistics). We find that the 1 σ constraints on S_8 do not vary more than 5 per cent across different NDEs. Although not shown here, we also repeated this test for all the other statistics (and combinations) considered in this work, and found differences below 5 per cent in all cases.

Appendix D: Additional likelihood tests

In this Appendix we perform extra tests on our estimated likelihoods using the `CosmoGridV1` simulations. First, for the 400 compressed data vectors at our disposal, we looked at the distribution of residuals for each entry of our data vector. This is shown in Fig. 18. The residuals are well described by a Gaussian, with no clear sign of strong deviations from Gaussianity. This is true also for the non-Gaussian statistics implemented in this work. As was already noted by [39], this is partially thanks to the data compression algorithm, which helps to give the compressed data a more Gaussian distribution due to the central limit theorem [51].

As a second test, we sample from the likelihood estimated using our NDEs at the `CosmoGridV1` cosmology, and compare with the distribution of the compressed data vector measured in the `CosmoGridV1` simulations. In particular, we sample the likelihood at $\Omega_\mathrm{m} = 0.26$, $S_8 = 0.26\sqrt{0.84/0.3}$, $w = -1$, and $A_\mathrm{IA} = 0$. For this test, we generated 400 new `CosmoGridV1` maps and we also marginalised over redshift uncertainties and multiplicative shear bias (in contrast to the `CosmoGridV1` maps used in the rest of the paper, where we fixed nuisance parameters to their mean values). Such a comparison is shown in Fig. 19, for the case of second moments and PWH S01+C01. Although not shown here, other sum-

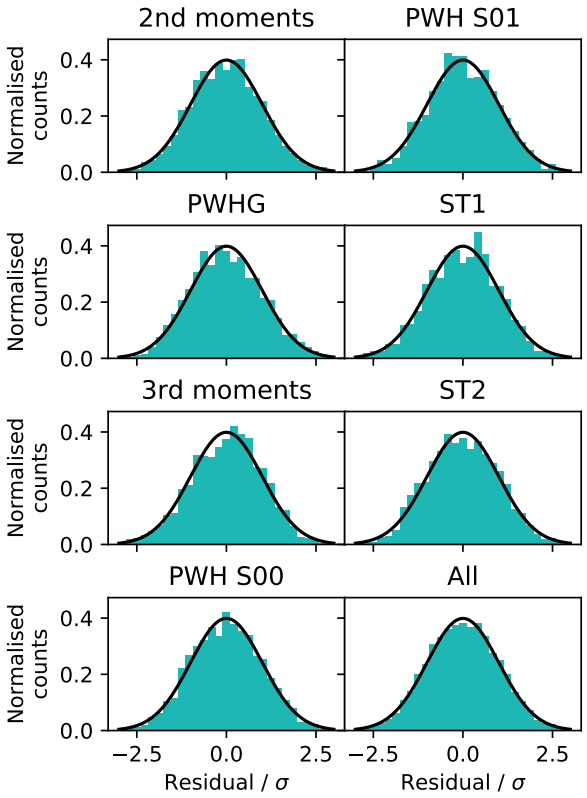

Figure 18. Residuals of individual data points in units of their expected standard deviation for the compressed data vector of the `CosmoGridV1` simulations. We compare to a Gaussian with zero mean and unit standard deviation.

mary statistics show a similar behaviour. The samples obtained from the NDEs match fairly well the distribution of compressed data vectors from the simulations, although for second moments they are slightly larger. This is expected: the likelihood estimated by the NDEs also marginalises over Ω_b, n_s, h_{100}, and neutrino mass, and we cannot fix them, because when training the NDEs we only made explicit the dependence on Ω_m, S_8, w, and A_IA. The `CosmoGridV1` samples do not marginalise over these additional parameters, so their distribution might be slightly smaller than the one predicted from the NDEs.

In Fig. 19 we also compare with the samples we would have obtained if we had assumed a Gaussian likelihood, estimating the mean and the covariance from the compressed `CosmoGridV1` measurements. These samples match very well the distribution of compressed measurements; this would not have been guaranteed had the likelihood been strongly non-Gaussian. Together with the residual tests (Fig. 18), this suggests that assuming a Gaussian likelihood for our compressed summary statistics could have been a reasonable option, at least at the `CosmoGridV1` cosmology. Of course, we cannot assume this generalises to other points in the parameter space, nor we could exclude *a priori* any cosmological dependence of the covariance. We note that our NDEs have learned that the likelihood is Gaussian at this point in parameter space, as the NDEs did not have any prior knowledge concerning the form of the likelihood.

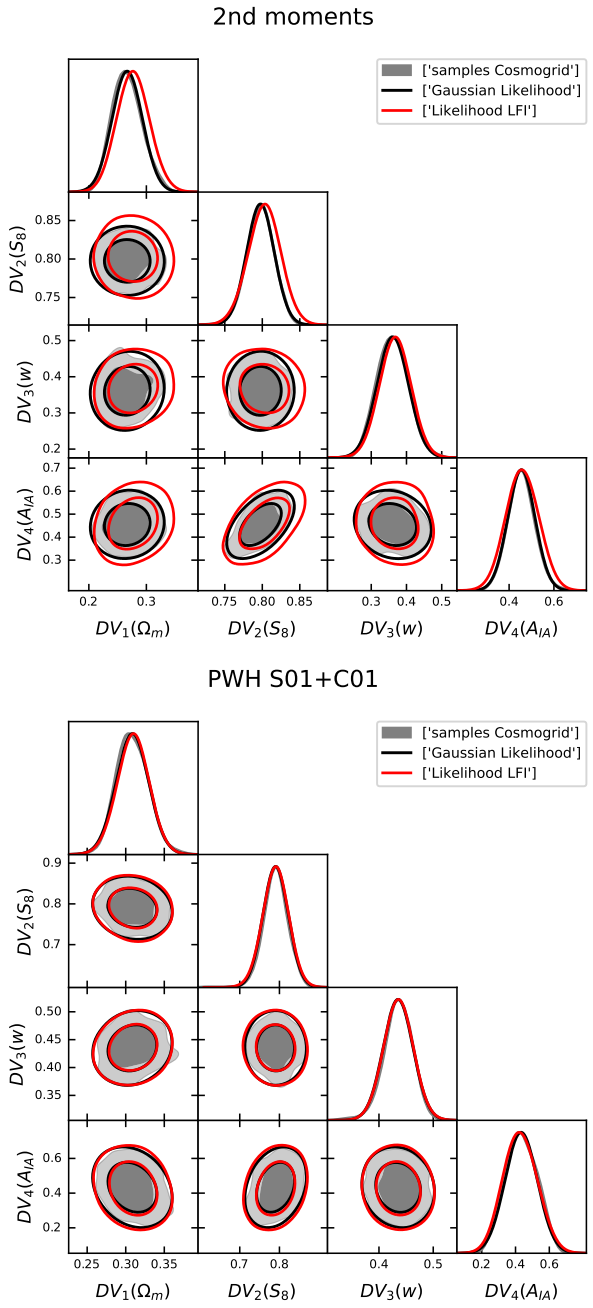

Figure 19. Samples of compressed summary statistics from the `CosmoGridV1` simulations (grey), compared to samples drawn from the learned likelihood at the `CosmoGridV1` cosmology (red). We also compare to the distribution we would have obtained had we assumed a Gaussian likelihood (black). The top panel refers to second moments and the bottom panel refers to PWH S01+C01.