Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **536,** 1303–1322 (2025) https://doi.org/10.1093/mnras/stae2629
Advance Access publication 2024 November 27

Dark energy survey year 3 results: likelihood-free, simulation-based *w*CDM inference with neural compression of weak-lensing map statistics

N. Jeffrey,[1]★ L. Whiteway,[1] M. Gatti,[2] J. Williamson,[1] J. Alsing,[3] A. Porredon,[4] J. Prat,[5,6,7] C. Doux,[2,8] B. Jain,[2] C. Chang,[6,7] T.-Y. Cheng,[5] T. Kacprzak,[9] P. Lemos,[5] A. Alarcon,[10,11] A. Amon,[12,13] K. Bechtol,[14] M. R. Becker,[10] G. M. Bernstein,[2] A. Campos,[15] A. Carnero Rosell,[16,17,18] R. Chen,[19] A. Choi,[20] J. DeRose,[21] A. Drlica-Wagner,[6,7,22] K. Eckert,[2] S. Everett,[23] A. Ferté,[24] D. Gruen,[25] R. A. Gruendl,[26,27] K. Herner,[22] M. Jarvis,[2] J. McCullough,[28] J. Myles,[29] A. Navarro-Alsina,[30] S. Pandey,[2] M. Raveri,[31] R. P. Rollins,[32] E. S. Rykoff,[28,24] C. Sánchez,[2] L. F. Secco,[7] I. Sevilla-Noarbe,[33] E. Sheldon,[34] T. Shin,[35] M. A. Troxel,[19] I. Tutusaus,[36] T. N. Varga,[37,38,39] B. Yanny,[22] B. Yin,[15] J. Zuntz,[40] M. Aguena,[17] S. S. Allam,[22] O. Alves,[41] D. Bacon,[42] S. Bocquet,[25] D. Brooks,[1] L. N. da Costa,[17] T. M. Davis,[43] J. De Vicente,[33] S. Desai,[44] H. T. Diehl,[22] I. Ferrero,[45] J. Frieman,[7,22] J. García-Bellido,[46] E. Gaztanaga,[11,42,47] G. Giannini,[7,48] G. Gutierrez,[22] S. R. Hinton,[43] D. L. Hollowood,[49] K. Honscheid,[50,51] D. Huterer,[41] D. J. James,[52] O. Lahav,[1] S. Lee,[23] J. L. Marshall,[53] J. Mena-Fernández,[54] R. Miquel,[55,48] A. Pieres,[17,56] A. A. Plazas Malagón,[28,24] A. Roodman,[28,24] M. Sako,[2] E. Sanchez,[33] D. Sanchez Cid,[33] M. Smith,[57] E. Suchyta,[58] M. E. C. Swanson,[26] G. Tarle,[41] D. L. Tucker,[22] N. Weaverdyck,[21,41] J. Weller,[38,39] P. Wiseman[57] and M. Yamamoto[18]

Affiliations are listed at the end of the paper

Accepted 2024 November 21. Received 2024 November 20; in original form 2024 March 4

ABSTRACT

We present simulation-based cosmological wcold dark matter (wCDM) inference using dark energy survey year 3 weak-lensing maps, via neural data compression of weak-lensing map summary statistics: power spectra, peak counts, and direct map-level compression/inference with convolutional neural networks (CNN). Using simulation-based inference, also known as likelihood-free or implicit inference, we use forward-modelled mock data to estimate posterior probability distributions of unknown parameters. This approach allows all statistical assumptions and uncertainties to be propagated through the forward-modelled mock data; these include sky masks, non-Gaussian shape noise, shape measurement bias, source galaxy clustering, photometric redshift uncertainty, intrinsic galaxy alignments, non-Gaussian density fields, neutrinos, and non-linear summary statistics. We include a series of tests to validate our inference results. This paper also describes the *Gower Street simulation suite*: 791 full-sky PKDGRAV3 dark matter simulations, with cosmological model parameters sampled with a mixed active-learning strategy, from which we construct over 3000 mock dark energy survey lensing data sets. For wCDM inference, for which we allow $-1 < w < -\frac{1}{3}$, our most constraining result uses power spectra combined with map-level (CNN) inference. Using gravitational lensing data only, this map-level combination gives $\Omega_{\rm m} = 0.283^{+0.020}_{-0.027}$, $S_8 = 0.804^{+0.025}_{-0.017}$, and $w < -0.80$ (with a 68 per cent credible interval); compared to the power spectrum inference, this is more than a factor of two improvement in dark energy parameter ($\Omega_{\rm DE}$, w) precision.

Key words: gravitational lensing: weak – cosmology: large-scale structure of Universe – cosmology: dark energy.

1 INTRODUCTION

Weak gravitational lensing induces a pattern in the observed shapes of galaxies; we may use this to infer the distribution of foreground matter, including visible matter and (invisible) dark matter. The lensing effect is sensitive both to large-scale structure formation and to geometric effects that probe the expansion history of the Universe.

Cosmological inference is typically performed using two-point correlation functions (e.g. power spectra) of the lensing signal. The currently most up-to-date analyses of this type are from the dark energy survey (DES; Amon et al. 2021; Secco et al. 2022), the Kilo-degree survey (KiDS; Asgari et al. 2021; Li et al. 2023b), and hyper

★ E-mail: n.jeffrey@ucl.ac.uk

suprime-cam (HSC; Li et al. 2023a). Two-point statistics capture only some of the cosmologically relevant information and so are limited in discovery potential, in comparison to the information encoded in the full lensing *mass map*; for DES Year 3 such lensing mass maps were presented in Jeffrey et al. (2021b).

This paper has two scientific aims: (i) to use map-level inference to better constrain the cosmological parameters of the 'w-cold-dark-matter' (wCDM) model, and (ii) to use simulation-based inference (also known as likelihood-free inference) methods to ensure realistic data modelling and reliable inference.

Deep learning methods (see Goodfellow, Bengio & Courville 2016 for an introduction) are used in two distinct ways in this analysis:

(i) Compression: We perform neural compression of high-dimensional data or summary statistics of the data; in our case we compress the map itself (using convolutional neural networks), the power spectra, and the peak counts from the map.

(ii) Neural likelihood estimation and validation: We use neural density estimation (as is typical with simulation-based inference) to learn the form of the likelihood from simulated mock data. We then validate the resulting posterior probability distributions.

This paper also serves as the public release of the *Gower Street simulation suite*, consisting of 791 (so far – the suite may grow in future) full-sky cosmological simulations that vary seven cosmological parameters of the wCDM model: the cosmological density parameter Ω_m, the amplitude parameter σ_8, the scalar spectral index n_s, the Hubble parameter $h = H_0/(100\ \mathrm{km\ s^{-1}\ Mpc^{-1}})$, the physical baryon density $\Omega_\mathrm{b}h^2$, the dark energy equation of state w, and the neutrino mass m_ν (the sum of the masses of the three neutrino mass eigenstates, quoted in electron volts). For the analysis in this paper, each full sky simulation can be split into four DES sky footprints, giving over 3000 quasi-independent mock DES surveys. Using multiple noise realizations, we augment this suite to over 10^4 non-independent mock DES surveys; these are used to train data compression and to perform simulation-based inference and posterior probability validation.

One novel aspect of this work is the combination of simulation-based inference and map-level inference for an application with state-of-the-art weak gravitational lensing data. Fluri et al. (2022) recently pioneered the use of deep learning for map-level weak lensing inference with KiDS data; this paper assumed a Gaussian likelihood. Other works have used machine learning methods to extract cosmological information, but without characterizing the likelihood with simulation-based inference (Fluri et al. 2018, 2019; Peel et al. 2019; Ribli et al. 2019). Jeffrey, Alsing & Lanusse (2021a) used both simulation-based inference and deep learning for cosmological feature extraction, but this work used only the DES science verification data and so did not produce a competitive cosmological result.

In a companion DES analysis, we are developing a simulation-based inference pipeline that uses wavelet scattering representations instead of convolutional neural networks, of which Gatti et al. (2024a) is an initial description. In further analyses, we will also try to understand the physical origin or environmental dependence of our map-level (deep learning) inference. These are all DES year 3 analyses, awaiting the final full DES year 6 data.

In Section 2, we introduce simulation-based inference, describing in turn the use of neural likelihood estimation to learn the form of the likelihood from realistic mock data (Section 2.2), the principle of data compression (Section 2.3), validation of the resulting posterior probability densities (Section 2.4), and parameter sampling and marginalization (Section 2.5).

In Section 3, we give an overview of weak gravitational lensing and in Section 4, we describe the Gower Street suite of simulations.

In Section 5, we describe the DES year 3 weak gravitational lensing data. We also describe how we generate mock DES data from the Gower Street simulations in a way that matches survey properties, noise, and forward modelling contributions to systematic uncertainty (e.g. intrinsic alignments of galaxies and photometric redshift uncertainty).

In Section 6, we describe each of the chosen summary statistics of the data and the data compression methods. The summary statistics described are the weak-lensing map itself (we describe how we construct convolutional neural networks to extract the cosmological information), the power spectra, and the counts of peaks in the lensing map.

We present the cosmological inference results in Section 7 and conclude in Section 8.

2 SIMULATION-BASED INFERENCE

2.1 Motivation

For parameter inference from complex physical systems, the likelihood, i.e. the conditional probability density $p(x|\theta)$ of the data x given the model parameters θ, is typically not known exactly or is too complex to be tractable. For these problems, *simulation-based inference* (also known as 'likelihood-free inference' or 'implicit inference') provides a solution.

For weak gravitational lensing data, the exact form of the likelihood is typically not known. This is due both to the non-linear evolution of the cosmological density field and to several complicated observational effects (survey masks, various systematic biases, non-Gaussian noise contributions, etc.). Even if we assume that the underlying density field is Gaussian, the two-point statistics in weak lensing can have a significantly non-Gaussian distribution, especially if realistic observational effects are included (Alsing, Heavens & Jaffe 2017; Sellentin & Heavens 2018; Sellentin, Heymans & Harnois-Déraps 2018; Taylor et al. 2019).

For higher-order statistics, there is typically no closed-form expression for the likelihood. Even the expectation values of $p(x|\theta)$ for many higher-order statistics (e.g. peak counts) must be estimated from simulated mock data. We cannot expect the probability density $p(x|\theta)$ for these statistics to be Gaussian, as there are multiple sources of non-Gaussianity in the data model.

Even if the likelihood were known to be Gaussian, for observables that used simulated predictions (e.g. peak counts or map-level deep learning) the covariance matrix also has to be estimated from a significant number of simulations, typically run with fixed input parameters. The simulation-based inference approach avoids this, and hence can still be highly applicable even in the Gaussian likelihood case.

Furthermore, even if the likelihood is known, simulation-based inference methods allow implicit marginalization over nuisance parameters. As discussed in Jeffrey & Wandelt (2020), traditional methods fail with large parameter spaces, whereas with simulation-based inference methods we can sidestep intractable high-dimensional inference and focus only on the selected parameters of interest. This implicit marginalization over nuisance parameters is central to the analyses presented in this paper, as we vary both unconstrained cosmological parameters and nuisance parameters (including $n(z)$

redshift distributions with $\sim 10^3$ dimensions). This is discussed further in Section 2.5.

2.2 Neural likelihood estimation

This work uses the *neural likelihood estimation* technique from the field of simulation-based inference; in this technique, the form of $p(x|\theta)$ is learned from mock data realizations (Alsing, Wandelt & Feeney 2018b; Papamakarios, Sterratt & Murray 2019).[1] By generating simulated mock data x_i, we are in fact drawing samples according to

$$x_i \sim p(x|\theta_i), \tag{1}$$

where θ_i are the input parameters to the simulation with index i. From a set of simulated mock data labelled by their parameter values $\{x_i, \theta_i\}$, we can then learn a density q that approximates the underlying probability density p, such that $p(x|\theta) \approx q(x|\theta)$.

In our case, θ is a chosen subset of the wCDM model parameters coupled with nuisance parameters corresponding to observational effects (e.g. intrinsic alignment amplitude).

Given parameters of interest θ and given some data x (e.g. the lensing map or its power spectrum), our first step is to estimate $p(x|\theta)$. This estimated likelihood is then evaluated for the observed data x_O, from which as usual the posterior probability density of the parameters can be related to the likelihood via Bayes' theorem:

$$p(\theta|x_O) = \frac{p(x_O|\theta)\ p(\theta)}{p(x_O)}\ . \tag{2}$$

To estimate the conditional distribution $p(x|\theta)$, we use the PYDELFI (Alsing et al. 2019) package[2] with an ensemble of neural density estimators (NDEs). NDEs use neural networks to parametrize densities, including (as here) *conditional* probability densities.

An NDE gives an estimate $q(x|\theta, \varphi)$ by varying the φ neural network parameters (e.g. weights and biases) to minimize the loss function

$$U(\varphi) = -\sum_{i=1}^{N} \log q(x_i|\theta_i; \varphi) \tag{3}$$

over the N forward-modelled mock data x_i. This loss corresponds to minimizing the Kullback–Leibler divergence (Kullback & Leibler 1951), a measure of change from the estimate q to the target p.

We have available two types of NDEs: Gaussian mixture density networks (MDN; Bishop 1994) and masked autoregressive flows (MAF; Papamakarios, Pavlakou & Murray 2017). An MDN represents the conditional density as a sum of several Gaussian components. A MAF is a type of normalizing flow i.e. it uses a series of bijective transformations from simple known densities (e.g. standard Gaussian) to the target density (Jimenez Rezende & Mohamed 2015; Kingma et al. 2016; Papamakarios et al. 2019).

For further details see Jeffrey et al. (2021a) (in which a similar neural likelihood estimation setup was used, and in which may be found a more technical introduction to these NDE methods).

However, unlike Jeffrey et al. (2021a), the results presented in this paper use only MAFs, as these were found to perform better at hard prior boundary edges (e.g. for $w \approx -1$). The MDNs were used only for validation with simulated data analyses. For the presented results (Section 7) we use an ensemble of four MAFs: each had either three, five, or six transformations (masked autoencoder for distribution estimation, i.e. MADE) with each using a neural network with two hidden layers (with widths of either 40 or 50).

2.3 Principle of data compression

Density estimation of $p(x|\theta)$ rapidly increases in difficulty as the dimensionality $\dim(x)$ of the data vector x increases (the 'curse of dimensionality'). In this DES weak-lensing analysis, the data dimensionality is $\sim 10^7$ for the case of map-level inference and $\sim 10^3$ for inference using power spectra and peak counts. Direct estimation of $p(x|\theta)$ is intractable.[3]

We take the (now standard) approach of data compression: apply some function $\mathcal{F}$ to the data to return compressed data $t = \mathcal{F}(x)$, while trying to preserve information about the parameters θ.

A poor compression (i.e. one that loses information) will not lead to biased inference. Because the same compression is applied consistently to both the simulated data and the observed data, a less-informative summary statistic $t_{\rm lossy}$ will lead to inflated posterior distributions on θ. In the limit of uninformative compression, any posterior distribution $p(\theta|t_{\rm lossy})$ will merely be equal to the prior $p(\theta)$.

Although we do not have to worry about poor compression leading to incorrect inference, we clearly want to find a compression scheme that is maximally informative with respect to the parameters of interest θ. Different techniques are available for compression, all of which aim to maximize the information content of t while dramatically reducing the dimensionality.

Under certain conditions it is possible to find $\mathcal{F}$ for which the dimension of t equals the number of inferred parameters, $\dim(t) = \dim(\theta)$, and which also is lossless with respect to the Fisher information (e.g. Heavens, Jimenez & Lahav 2000; Alsing & Wandelt 2018).

Neural compression, which we use in this DES analysis, takes advantage of the flexibility of neural networks to parametrize $\mathcal{F}$. The neural network is trained using simulated mock data. Existing methods include the information maximizing neural network (Charnock, Lavaux & Wandelt 2018), which maximizes the Fisher information, and variational mutual information maximization (VMIM; Jeffrey et al. 2021a), which maximizes the mutual information between the compressed data and the target parameters.

Instead of these methods, we use a mean-square error (MSE) loss function to compress the data. This corresponds to an estimate of the mean of the posterior distribution for each parameter. Such a point estimate is clearly informative about the target parameters, and can be contrasted with the maximum likelihood parameter estimate, which corresponds to an optimal score compression (with some caveats: Alsing & Wandelt 2018). We do not expect this MSE compression to be optimal (e.g. compared to VMIM), but it is simple to implement.

The network architecture for the compression used in this work is described in detail in Section 6. As the MSE only depends on the marginal posterior per parameter, we train a different network per parameter. Multiple noise realizations serve as data augmentation in our training data for compression. Throughout this analysis, the neural compression is learned from different noise realizations of the

[1] c.f. Cranmer, Pavez & Louppe (2015) introducing of neural likelihood ratios estimation.

[2] https://github.com/justinalsing/pydelfi

[3] Recent innovations may provide an alternative approaches (e.g. Kingma et al. 2024) without compression in future.

mock data to those that are used for neural likelihood estimation—this is to avoid overfitting.

2.4 Posterior probability validation

2.4.1 Coverage tests

Coverage tests in Bayesian analysis check whether credible intervals have the expected probabilities. Looking at one-dimensional marginalized posteriors, we define a particular *credible interval* to be the narrowest interval containing (say) 90 per cent of the probability weight; other credible intervals would work equally well (this can be generalized e.g. Lemos et al. 2023). View the inference process as a procedure which, given observed data x_O, yields a posterior distribution $p(\theta|x_O)$ and hence a credible interval for θ. In the coverage test we use a parameter $\theta_{\rm test}$, selected from the prior $p(\theta)$, as input to a simulation yielding output data $x_{\rm test}$, from which we derive a posterior $p(\theta|x_{\rm test})$ and hence a credible interval; if the inference process is correctly implemented then the true test parameter value $\theta_{\rm test}$ will fall in this credible interval 90 per cent of the time. By repeating with many such $\theta_{\rm test}$, we are able to gain confidence that our estimated posterior distributions are indeed correct (Prangle et al. 2014; Hermans et al. 2021).

This test is relatively straightforward for this type of simulation-based inference, for which we have a number of existing mock data simulations and where the inference scheme is *amortized* (and so fast to evaluate probabilities for new data). Coverage testing is a useful aspect of inference, ensuring that the results are reliable, which is often unfeasible with traditional statistical approaches.

Given the computational expense of each simulation giving us a limited supply of mock data realizations, the biggest risk of failure is that we have insufficient simulations to robustly estimate the likelihood. Coverage tests can reassure us that we have sufficient numbers of simulations for this task; a successful coverage test implies there were enough simulations to accurately estimate the likelihood.

In this analysis, we show successful coverage tests for inference using our learned likelihoods; this serves as one validation of the posterior distribution obtained for the actual observed data $p(\theta|x_O)$.

2.4.2 Neural density ensemble convergence

The individual likelihood estimates from the neural density ensemble can be used as a further validation step. The individual density estimates will converge to a common value as the number of simulations increases; therefore, if the posterior distributions from each independent density estimation are in disagreement, this would be evidence that we had an insufficient number of simulated mock data realizations.

2.5 Parameter sampling and marginalization

The main strength of neural *likelihood* estimation (learning $p(x|\theta)$) rather than the neural *posterior* estimation (learning $p(\theta|x)$) is that the parameters θ in the training data (the simulations) do not have to be drawn from the prior $p(\theta)$.

This has two benefits. The first is that the prior can be changed at will after the simulations have been run; for example, it is possible to take new external information into account (e.g. by simply combining likelihoods). The second, of particular importance to this work, is that additional simulations can be run in regions of parameter space that are most useful for the neural density estimation; this is known as *active learning*. One can choose the parameter values for the new simulation from some *acquisition function*, which may be based on the existing posterior estimates, to improve robustness. In this DES analysis, this was implemented in two stages: (1) most σ_8 and $\Omega_{\rm m}$ parameters were at first distributed according to the existing DES analysis constraints, and (2) 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 high-NDE ensemble variance (see Section 2.4.2). Our sampling scheme is discussed further in Section 4.

For the parameters that are not part of the active learning scheme, we can still choose to distribute them according to a prior. If any set of parameters $\theta_{\rm marginal}$ is distributed according to the chosen prior $p(\theta_{\rm marginal})$, and if these parameters are excluded from the parameter set θ used for neural likelihood estimation, then these $\theta_{\rm marginal}$ parameters will be implicitly marginalized during inference. This is explained via *marginal posterior density* estimation in Jeffrey & Wandelt (2020). The uncertainty in these parameters $\theta_{\rm marginal}$ is still accounted for in the resulting posterior distributions, as the parameters are varied in each simulated mock data realization, but the parameters are implicitly marginalized, which avoids explicit (and intractable) high-dimensional density estimation and unnecessary marginalization integration.

In the Gower Street simulations, all parameters other than σ_8 and $\Omega_{\rm m}$ are drawn from their prior distributions (with some caveats; Section 4). All observational nuisance parameters (e.g. intrinsic alignment and redshift) are also drawn from their priors in the mock DES lensing map generation (Section 5). This allows implicit marginalization if necessary.

3 WEAK GRAVITATIONAL LENSING

Weak gravitational lensing (WL; Bartelmann & Schneider 2001) is the coherent slight alteration (for us, primarily *shearing*) of the shapes of distant 'source' galaxies by the gravitational influence of intervening matter (mostly dark). The unlensed shapes are unknown, and thus act as a noise term in WL analysis; the large surface density of galaxies visible in modern surveys allows the WL convergence signal (a weighted average of the overdensity along the line-of-sight, convolved with the redshift density of source galaxies) to be measured despite this noise. The two-point correlation functions of the signal may be estimated from observed data and compared to theoretical predictions from a cosmological model, thereby constraining the parameters of the model. Alternatively, convergence *maps* may be constructed. The convergence field is not a Gaussian random field, and so the power spectrum is not the full story; there is further information in various beyond-two-point statistics from these maps. Alas, theoretical model predictions for these so-called *non-Gaussian* statistics are generally not available; results derived from simulations must be used instead.

This section describes briefly the theory required to link the WL shear observables to results obtained from theory or simulations.

3.1 Theory

We follow Jeffrey et al. (2021b) section 2; see that paper for full details. See Fig. 1 for a schematic diagram of the relationships between the fields discussed.

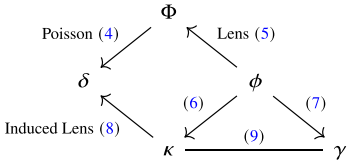

Figure 1. The relationships among weak-lensing fields Φ (gravitational potential), δ (overdensity), ϕ (weak-lensing potential), κ (convergence), and γ (shear); these relationships allow us to link observations to cosmological theory and simulations. Arrows represent spatial second derivatives; the line between κ and γ is a relationship of harmonic coefficients. Numbers refer to the corresponding equations in Section 3.1, where further details are given.

The gravitational potential Φ and the matter overdensity field $\delta \equiv \rho/\bar{\rho} - 1$ are related by the Poisson equation

$$\nabla_r^2 \Phi(t, \boldsymbol{r}) = \frac{3\Omega_\mathrm{m} H_0^2}{2a(t)} \delta(t, \boldsymbol{r}) \,. \tag{4}$$

Here, $\boldsymbol{r}$ is a comoving spatial coordinate and a is the scale factor.

The weak-lensing potential ϕ is defined via the lens equation; ϕ is sourced by the gravitational potential, together with a lensing efficiency factor (written here assuming a flat Universe), all integrated along the line of sight to a source galaxy at comoving distance χ (here we use the Born approximation), and then further integrated over the redshift distribution $n(z)$ of source galaxies:

$$\phi(\theta, \varphi) = \frac{2}{c^2} \int_0^\infty \mathrm{d}\chi \; n(z(\chi)) \int_0^\chi \mathrm{d}\chi' \frac{(\chi - \chi')}{\chi \chi'} \Phi(\chi', \theta, \varphi). \tag{5}$$

The weak lensing potential is defined on the celestial sphere, so it is convenient to use the formalism of spin-weight functions on the sphere; see Castro, Heavens & Kitching (2005) for details and see also Sellentin et al. (2023) appendix A for geometrical comments. Let $\eth$ and $\bar{\eth}$ denote the spin-weight covariant derivative and its adjoint. Let κ and γ be the weak-lensing convergence (spin-weight 0) and shear (spin-weight 2); they are second derivatives of the weak-lensing potential:

$$\kappa = \frac{1}{4}(\eth\bar{\eth} + \bar{\eth}\eth)\phi \tag{6}$$

$$\text{and} \quad \gamma = \frac{1}{2}\eth\eth\phi. \tag{7}$$

Equations (4), (5), and (6) yield an induced lens equation linking δ and κ:

$$\kappa(\theta, \phi) = \frac{3\Omega_\mathrm{m} H_0^2}{2c^2} \int_0^\infty \mathrm{d}\chi n(z(\chi)) \int_0^\chi \mathrm{d}\chi' \frac{\chi'(\chi - \chi')}{\chi} \frac{\delta(\chi', \theta, \phi)}{a(\chi')}. \tag{8}$$

Finally we move to harmonic space, representing an arbitrary field b by its coefficients $b_{\ell m}$ with respect to the basis of spherical harmonic functions of the appropriate spin-weight. Now $\eth$ and $\bar{\eth}$ behave in a simple fashion in this basis, and so equations (6) and (7) yield

$$\gamma_{\ell m} = -\sqrt{\frac{(\ell - 1)(\ell + 2)}{\ell(\ell + 1)}} \kappa_{\ell m}. \tag{9}$$

Using κ as a link, equations (8) and (9) together connect γ (which may be measured using the shapes of source galaxies – taking into account shape noise and other sources of noise, partial sky coverage, and systematic effects such as intrinsic alignments) with δ (which may be treated theoretically – at least up to two-point statistics – or modelled via simulations). We thus have the desired link between theory (or simulations) and observations.

4 GOWER STREET SIMULATIONS

4.1 Simulation configuration

The Gower Street suite of simulations consists of 791 gravity-only full-sky N-body simulations, produced using the PKDGRAV3 code (Potter, Stadel & Teyssier 2017), spanning a seven-dimensional parameter space in wCDM (Ω_m, σ_8, n_s, h, $\Omega_\mathrm{b}h^2$, w, m_ν).

For reviews of the theory of simulations, see Efstathiou et al. (1985) and Angulo & Hahn (2022). In common with other N-body simulation codes, PKDGRAV3 uses a box of side L, filled with N^3 particles. At a start time, corresponding to redshift z_0, the particles are arranged in phase space (positions and velocities) so as to match desired initial conditions; for this, PKDGRAV3 uses second-order Lagrangian perturbation theory. The positions/velocities of the particles are then updated (under the influence of gravity – modelled as Newtonian – against the backdrop of Universe expanding according to specified cosmological parameters) to yield snapshots of positions/velocities at various discrete times.

From this four-dimensional data set, PKDGRAV3 extracts *lightcone* data i.e. it restricts the data set to events currently visible to the theoretical observer at the centre of the simulation. Specifically, the code estimates each particle's worldline (by interpolating between the particle's known positions at each time slice) and calculates, in four-dimensional space, the intersection of this worldline (a one-dimensional curve) with the observer's lightcone (a three-dimensional cone). This gives, for each particle, what event (redshift and position) on its worldline is currently visible. These data are then binned, by redshift (a bin corresponds to the redshift interval between two snapshots) and by position on the sky (into HEALPIX, Górski et al. 2005 pixels). The results (particle count per pixel per redshift bin) are then output, with one file per redshift bin. For higher redshifts, the comoving distance to the redshift will exceed the box side L; to avoid this, the simulation box is replicated M times in each direction (a total of $(2M)^3$ replications, with the observer at the centre of this 'superbox').

The Gower Street simulations use $L = 1250\,h^{-1}$ Mpc and $N = 1080$. We set the initial redshift to $z_0 = 49$, and we produce 101 snapshots (and hence 100 lightcone files), equally spaced in proper time between z_0 and redshift zero. For the HEALPIX pixelization, we set NSIDE $= 2048$. The simulation box is replicated $M = 10$ times,[4] although the bulk of our redshift distributions ($z < 1.5$) can be covered by only three replications.

Fig. 2 presents an example of such a simulation; the map shows the matter overdensity as derived from the PKDGRAV3 particle count.

To validate the output of our simulations, we saved the three-dimensional particle positions as a final redshift snapshot at $z = 0$ for a single simulation run whose parameters were $\Omega_\mathrm{m} = 0.3001$, $\sigma_8 = 0.7894$, $n_s = 0.95$, $h = 0.687$, $\Omega_\mathrm{b}h^2 = 0.022\,43$, $w = -0.95$, $m_\nu = 0.065$. Limits on computation time and disc space prevented us from generating these for multiple simulations. From this snapshot, we measured the matter power spectrum using the NBODYKIT (Hand et al. 2018) code. Fig. 3 compares (a) the measured power spectrum from this simulation to (b) the theoretical power spectrum calculated using the *Euclid* emulator (Knabenhans et al. 2021) code. At small scales, where the finite resolution of the simulation would be expected to cause inaccuracies, the difference between the measured and theoretical power spectrum remains below 2 per cent. This is

[4]We thank Janis Fluri for code amendments allowing an increase from the default $M = 3$ value.

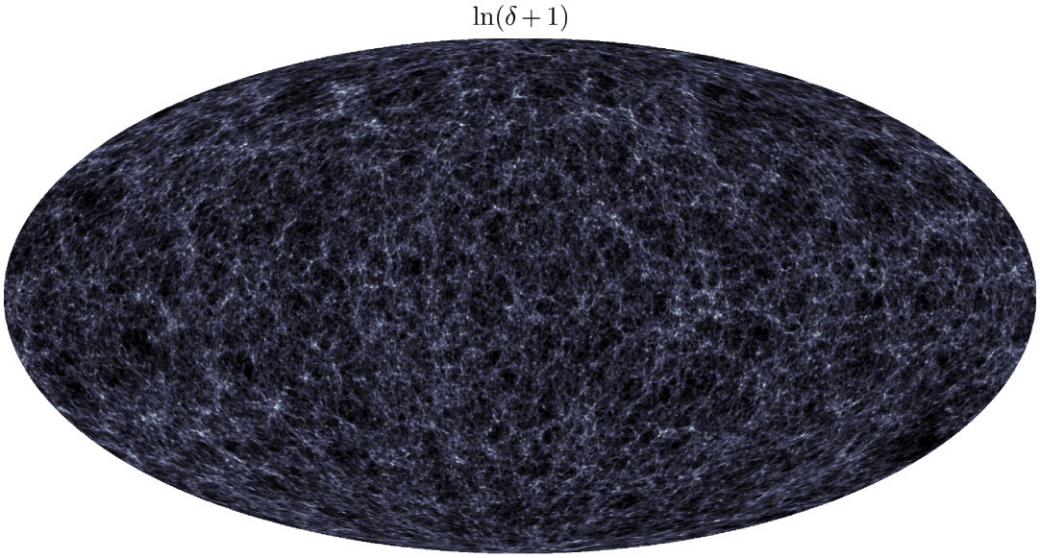

Figure 2. Example dark matter simulation from the Gower Street simulation suite. This map on the celestial sphere (Mollweide projection) uses the average overdensity δ from all shells up to a redshift $z = 0.15$. Such simulations form the basis of the mock DES Y3 weak lensing maps used in the inference pipeline.

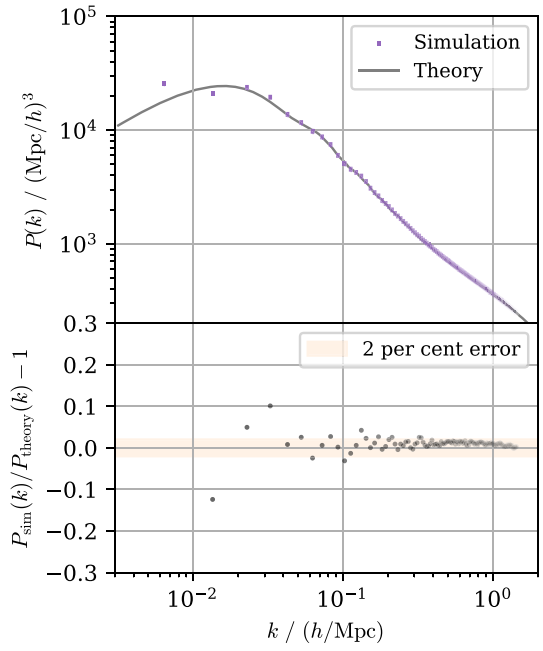

Figure 3. Comparing the matter power spectrum $P(k, z = 0)$ of a simulation to that from theory. The theory prediction for the power spectrum $P_{\rm theory}(k)$ combines CAMB for linear theory (Lewis, Challinor & Lasenby 2000) and the *Euclid* emulator (Knabenhans et al. 2021) for the non-linear contribution. At small scales, where the finite resolution of the simulation is expected to cause inaccuracies, the systematic error remains below 2 per cent.

within the relative error between different non-linear power spectrum prescriptions and other modelling choices, such as choice of neutrino model or astrophysical feedback model. Baryon feedback effects are not included in the simulation suite, but their effects are tested in Section 7.

This paper serves as the formal release of these simulations, which are available at www.star.ucl.ac.uk/GowerStreetSims/.

4.2 Cosmological parameters

A total of 791 simulations were performed. The first 192 of these were 'verification' runs, done to test the software pipeline; they had a naive handling of neutrinos. For these runs, the initial conditions were specified to PKDGRAV3 via a transfer function generated using NBODYKIT (Hand et al. 2018); in all of these runs the neutrino mass was fixed to 0.06. Neutrinos played no role beyond this in these initial simulations. Further simulations, beyond these initial verification runs, had a more sophisticated handling of neutrinos, done via the CONCEPT software (see Tram et al. 2019).

Each simulation was given its own values for seven cosmological parameters within wCDM: $\Omega_{\rm m}$, σ_8, n_s, h, $\Omega_{\rm b}h^2$, w, and m_ν; in addition, each simulation had a different value for the random seed used when generating initial conditions (so that the simulations also display a range of behaviours arising from cosmic variance). The two parameters for which weak-lensing observations are most constraining, $\Omega_{\rm m}$ and σ_8, were chosen via *active learning*: During later runs, values for these parameters were chosen so to maximize the utility of the simulation suite (i.e. concentrated in regions of parameter space that were both important and underrepresented), by minimizing the variance between the NDEs in the ensemble. For simplicity, for the initial simulations the parameters were sampled from the posterior distribution of the parameters (both from the existing published DES results and as calculated using the simulations so far – see Alsing, Wandelt & Feeney 2018a), but with a hard exclusion zone around already-used parameter combinations.

The remaining parameters were chosen (independently) as follows:

(i) $n_{\rm s} \sim \mathcal{N}(0.9649, 0.0063)$; from Planck (Planck Collaboration 2021) but with the standard deviation boosted by a factor of 1.5.

(ii) $h \sim \mathcal{N}(0.7022, 0.0245)$; consistent with SH0ES (Riess et al. 2022) and Planck (Planck Collaboration 2021), as its mean is midway between the means of these experiments, and its one standard deviation contour encompasses the two standard deviation contours of both experiments.

(iii) $\Omega_{\rm b}h^2 \sim \mathcal{N}(0.022\,37, 0.000\,15)$; from Planck (Planck Collaboration 2021).

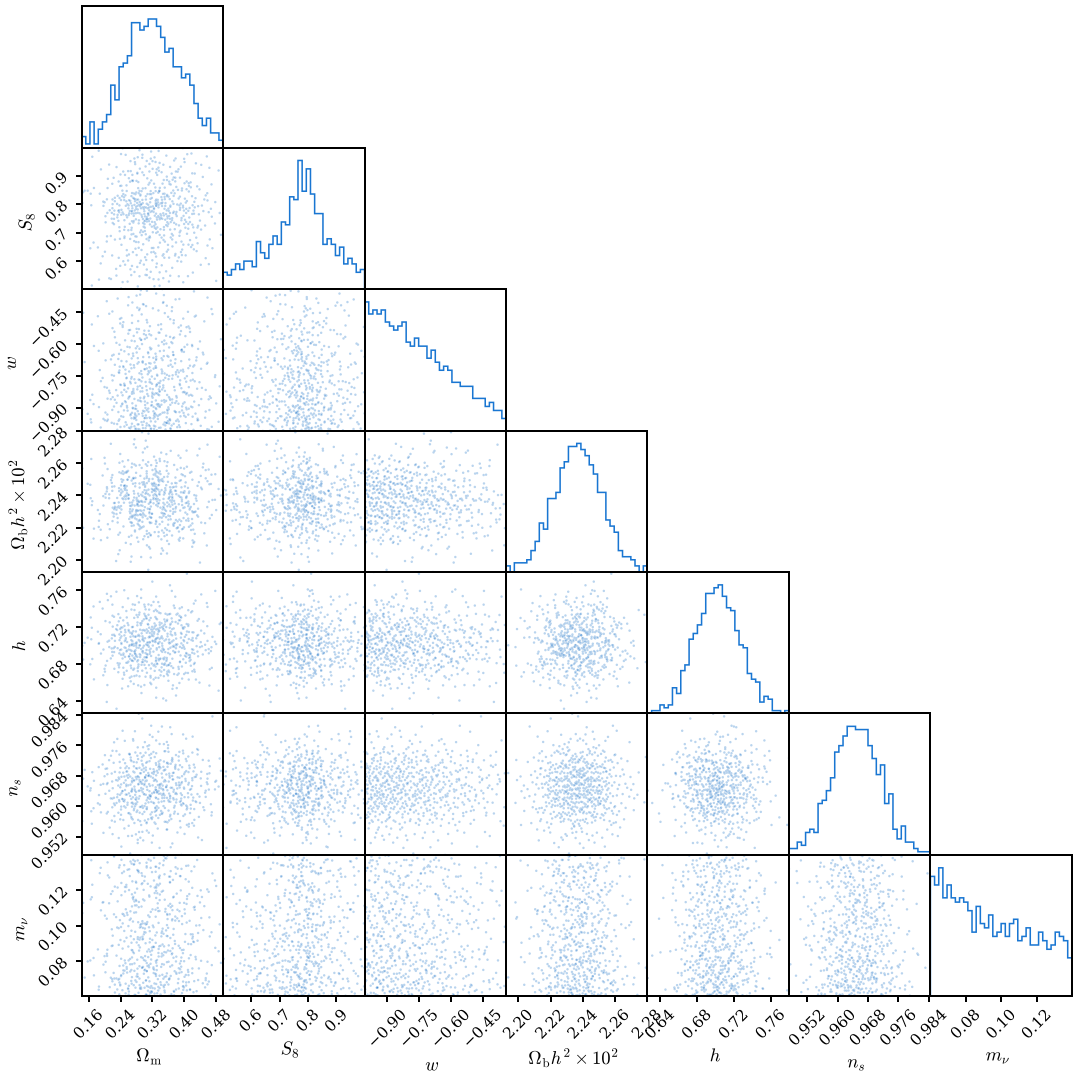

Figure 4. Parameter values used in the Gower Street simulations. The cosmological parameters span variations the in νwCDM model. We exclude the initial 192 'verification' runs to simplify the presentation of the neutrino mass distribution. For all parameters other than Ω_m and S_8, these parameters are distributed according to their prior probability distributions. For Ω_m and S_8, we always use neural likelihood estimation to condition on these parameters, removing the dependence on their simulated distribution.

(iv) $w \sim \mathcal{N}(-1, 1/3)$, but with values less than -1 or greater than $-1/3$ then discarded. However, for the first 128 runs (part of the 'science verification' runs), this discarding was not done (resulting in approximately 64 runs with $w < -1$). These runs have been kept as they help to smooth what would otherwise be the discontinuity at $w = -1$.
This choice of $w > -1$ excludes phantom dark energy. This has some theoretical justification for an N-body simulation on an expanding background, but is also motivated by computational limitations with low values of Ω_m when using PKDGRAV3 and CONCEPT.

(v) m_ν: As described in more detail above, fixed at 0.06 for the initial 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.

The parameter values are not sampled through i.i.d. draws from their respective distributions. Instead, to avoid similar parameter combinations arising due to random chance, we sample a multivariate uniform distribution using a mixture of Sobol and Halton sequences (and then transform where necessary from uniform to Gaussian by applying the inverse function of the Gaussian cumulative distribution function).

Fig. 4 displays the parameter values for pairs of parameters used in the Gower Street simulations. We only show the parameter combinations for the simulations not included in the initial 192 'verification' runs, so that we can include the neutrino mass in the figure.

5 DARK ENERGY SURVEY DATA AND MOCK DATA MODELLING

5.1 DES Year 3 weak lensing data

DES is a photometric galaxy survey that covers ~ 5000 deg^2 of the South Galactic cap. Mounted on the Cerro Tololo Inter-American Observatory four metre Blanco telescope in Chile, the 570 megapixel

Dark Energy Camera (Flaugher et al. 2015) images the field in $griz Y$ filters. We use data from the first three years of the survey (DES Y3).

The simulated galaxy catalogues are created so as to match DES Y3 for *known* properties. For example, the sky mask is known but the intrinsic alignment model parameters are not, so we simulate with fixed sky mask but vary the intrinsic alignment amplitude in each simulation.

The DES Y3 shear catalogue (Gatti et al. 2021), built upon the Y3 Gold catalogue (Sevilla-Noarbe et al. 2021), uses the METACALIBRATION algorithm (Huff & Mandelbaum 2017; Sheldon & Huff 2017) to measure galaxy ellipticities from noisy images. The raw images were processed by the DES Data Management team (Sevilla et al. 2011; Abbott et al. 2018; Morganson et al. 2018).

METACALIBRATION provides an estimate of the shear field using a self-calibration framework that uses the data itself to correct for selection effects in the response of the estimate to shear. Inverse variance weights are assigned to galaxies. The DES Y3 shear catalogue has 100 204 026 objects, with a weighted $n_{\rm eff} = 5.59$ galaxies arcmin^{-2}. The METACALIBRATION self-correction accounts for most of the multiplicative bias, but there is a remaining multiplicative bias of 2 per cent to 3 per cent (MacCrann et al. 2022). This multiplicative factor is left uncalibrated but is parametrized and its uncertainty accounted for in our inference framework.

The shear catalogue has also been tested for additive biases (e.g. due to point spread function residuals; see Gatti et al. 2021). The catalogue is characterized by a non-zero mean shear which is subtracted at the catalogue level before performing any analysis. The shear catalogue is divided into four tomographic bins, selected so as to have roughly equal number density.

The catalogue is used to create shear maps with a HEALPIX pixelization of NSIDE = 512. This relatively low resolution removes small scales that we cannot confidently model. This is tested and discussed further in Section 7. The estimated value of the shear field in the map pixels is given by:

$$\gamma^{\nu}_{\rm obs} = \frac{\sum_j \epsilon^{\nu}_j w_j}{\bar{R} \sum_j w_j}, \quad \nu = 1, 2, \tag{10}$$

where ν refers to the two shear field components, w_j is the per-galaxy inverse variance weight, $\bar{R}$ is the average METACALIBRATION response of the sample, and the summations are taken over the galaxies lying in a particular pixel.

5.2 Simulation map raytracing

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. The lens planes are provided as HEALPIX maps and are obtained from the raw number particle counts:

$$\delta_{\rm shell}(\phi, {\rm s}) = \frac{n_{\rm part}(\phi, {\rm s})}{\langle n_{\rm part}(\phi, {\rm s})\rangle_\phi} - 1 \ , \tag{11}$$

where $n_{\rm part}(\phi, {\rm s})$ is the number of particles in pixel ϕ for shell s and $\langle\rangle_\phi$ denotes an average over pixels.

The lens planes are converted into convergence planes $\kappa_{\rm shell}(\phi, \chi)$ under the Born approximation using the BORNRAYTRACE code.[5] The shear planes $\gamma_{\rm shell}(\hat{\mathbf{n}}, \chi)$ are obtained from the convergence maps using equation (9), the inverse Kaiser & Squires (1993) algorithm. We down-sample from the original resolution of NSIDE = 2048 to NSIDE = 512 (with pixel size ≈ 7.2 arcmin).

[5] https://github.com/NiallJeffrey/BornRaytrace

These convergence κ and shear γ maps are the true shear and convergence fields in thin redshift shells. To generate mock lensing maps as they would be observed, we must (a) integrate over a mock redshift distribution $n(z)$ (see equation 8), (b) simulate the effect of intrinsic alignment of galaxies, and (c) add the effect of galaxy shape noise and missing data (i.e. sky masks).

5.3 Intrinsic alignments of galaxies

We model the intrinsic alignment of galaxies using a density-weighted non-linear alignment (NLA) model.

Using the NLA model (Hirata & Seljak 2004; Bridle & King 2007), we relate the convergence signal that would result from pure intrinsic alignments (with no lensing), $\kappa_{\rm IA}$, linearly to the local density field:

$$\kappa_{\rm IA}(\phi, z) = -A_{\rm IA} C_1 \rho_{\rm crit} \frac{\Omega_M}{D(z)} \left(\frac{1+z}{1+z_0}\right)^{\eta_{\rm IA}} \delta(\phi, z) \tag{12}$$

in a pixel ϕ and for some shell redshift z. We use the standard value of $z_0 = 0.62$ and set $C_1 = 5 \times 10^{-14}\, {\rm M}_\odot\, h^{-2}\, {\rm Mpc}^2$ (as per Bridle & King 2007).

The density-weighting in our forward model modulates the standard NLA model, because the source galaxies trace the underlying density field, and so are preferentially observed in higher density regions. This effect is the same as the clustering term in the tidal-torque alignment model for intrinsic alignments (Blazek et al. 2019). The implementation of the source clustering effect is discussed in Section 5.4.

The amplitude of intrinsic alignments $A_{\rm IA}$ and the redshift evolution parameter $\eta_{\rm IA}$ are allowed to vary in our analysis as nuisance parameters. By sampling each of these parameters from a prior, they will be implicitly marginalized as part of our simulation-based inference procedure (see Section 2.5). We choose the following (weakly informative) priors for these parameters:

(i) $A_{\rm IA} \sim \mathcal{U}(-3, 3)$.
(ii) $\eta_{\rm IA} \sim \mathcal{U}(-5, 5)$.

The $\kappa_{\rm IA}$ maps are generated with the BORNRAYTRACE code using the simulated overdensity maps δ. From these we generate shear maps that contain only intrinsic alignment signal (i.e. no lensing).

5.4 Realistic mock shear maps

5.4.1 Source clustering

Due to the effect of clustering of source galaxies, known as *source clustering* (described and detected in Gatti et al. 2024b), it would be insufficient to assume a single galaxy redshift distribution $n(z)$ that is constant across the sky. Instead, our model $n(z, \phi)$ of the galaxy redshift distribution depends on sky position via an input HEALPIX pixel ϕ.

When constructing shear maps from observed data catalogues, each HEALPIX pixel is assigned an average shear, the average taken over all galaxies that are within that pixel and that are in the correct tomographic redshift bin. Since these source galaxies trace the underlying large-scale structure, higher-order correlations between the number of galaxies in pixels and the weak lensing signal encoded in the shear become important.

In our forward model for generating mock data, we model a per-pixel redshift distribution via the sky-averaged redshift distribution $\bar{n}(z)$, then modulated by the density of galaxies:

$$n(z) \propto \bar{n}(z)(1 + b_g \delta). \tag{13}$$

The modulation factor assumes a linear galaxy biasing model with bias parameter b_g; here δ is the matter overdensity as before. This modulation is combined with an overall rescaling of the shape noise contribution to preserve the expected overall noise variance. We follow the procedure of Gatti et al. (2024b), which contains further details.

We use a fiducial value of $b_g = 1$ throughout. Our tests find that changing this value does not significantly change our results, which we show in Section 7.3. The level of sensitivity to the biasing value, implies that linear bias is sufficient for this DES analysis (noting that this source clustering effect is typically not included in weak lensing analyses), but the effect of non-linear galaxy bias may become significant for future surveys.

5.4.2 *Shape noise and mask*

This procedure also uses the randomly rotated shapes of the observed DES catalogue galaxies to implicitly generate the average intrinsic shapes in our mock observations, contributing shape noise to our mock shear maps.

The sky mask does not have to be treated separately; it is simply the set of pixels that contain no source galaxies. Because the same sky mask is present in both the mock simulated data and the observed DES data, it will be implicitly taken into account as part of our inference pipeline.

5.4.3 *Multiplicative shear bias*

To account for the residual errors in the shape measurement, we include a multiplicative shear bias in the forward model of the mock data. For a multiplicative shear bias m, associated with the particular tomographic bin, we rescale the shear by a factor of $1 + m$.

Using the results from image simulations as presented in MacCrann et al. (2022), we use the following priors on m for the various tomographic bins:

(i) $m_1 \sim \mathcal{N}(-0.0063, 0.0091)$.
(ii) $m_2 \sim \mathcal{N}(-0.0198, 0.0078)$.
(iii) $m_3 \sim \mathcal{N}(-0.0241, 0.0076)$.
(iv) $m_4 \sim \mathcal{N}(-0.0369, 0.0076)$.

5.4.4 *Photometric redshift uncertainty*

In the above discussion of source clustering, we used a fixed sky-averaged redshift distribution $\bar{n}(z)$. In fact, we sample realizations of possible distributions that are consistent with the data, based on the HYPERRANK methodology (see Cordero et al. 2021) using photometric redshift data $x_{\rm phot}$.

The four tomographic bins are constructed to have roughly equal number density (Myles et al. 2021) and the initial redshift distributions are provided by the SOMPZ method (Myles et al. 2021) in combination with clustering redshift constraints (Gatti et al. 2022) and correction due to the redshift-dependent effects of blending (MacCrann et al. 2022). Rather than using the best-guess $\bar{n}(z)$, HYPERRANK generates realizations of possible $\bar{n}(z)$ samples in a way that marginalizes over redshift uncertainty.

Fig. 5 shows a random selection of $\bar{n}(z)$ samples. Each mock realization uses a different randomly sampled $\bar{n}(z)$; this contributes uncertainty in the photometric redshift distributions through our forward model, so that this uncertainty is taken into account in the inference pipeline.

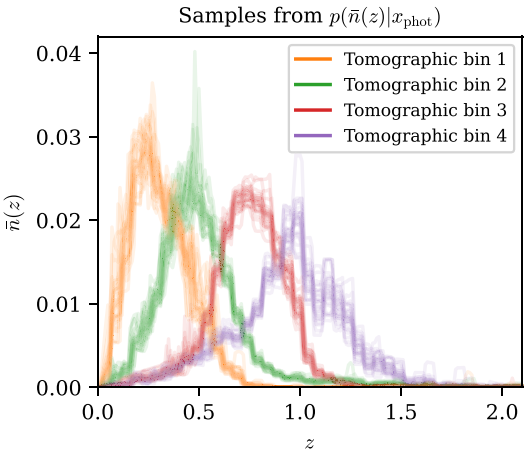

Figure 5. An example set of samples from the HYPERRANK probability distribution $p(\bar{n}(z)|x_{\rm phot})$ of the sky-averaged redshift distribution $\bar{n}(z)$ given the data used by HYPERRANK.

5.4.5 *Mock shear map summary*

In summary (following Gatti et al. 2024b), the mock shear signal at HEALPIX pixel ϕ in a thin simulated shell labelled with its redshift z is generated according to

$$\gamma(\phi) = \frac{\sum_z \bar{n}(z)[1 + b_g\delta(\phi, z)](1 + m)[\gamma(\phi, z) + \gamma_{\rm IA}(\phi, z)]}{\sum_z \bar{n}(z)[1 + b_g\delta(\phi, z)]} + \left(\frac{\sum_z \bar{n}(z)}{\sum_z \bar{n}(z)\left[1 + b_g\delta(\phi, z)\right]}\right)^{1/2} F(\phi)\,\frac{\sum_g w_g e_g}{\sum_g w_g}. \quad (14)$$

where $\bar{n}(z)$ is a HYPERRANK sample that varies between each mock simulation.

The $F(\phi)$ factor provides the overall rescaling of the noise; we use $F(\phi) = A(1 - B\sigma_e^2(\phi))^{1/2}$ where $A = [0.97, 0.985, 0.990, 0.995]$ and $B = [0.1, 0.05, 0.035, 0.035]$ for the four tomographic bins, and where $\sigma_e^2(\phi)$ is the shape noise pixel variance.

6 SUMMARY STATISTICS AND COMPRESSION

6.1 Map making and scale cuts

The mock data is prepared using DES Y3 footprints in HEALPIX format, as described in 5.4. These shear maps are degraded to NSIDE = 512, corresponding to a scale cut of 6.9 arcmin. Such a hard cut in pixel space corresponds to a smooth suppression of power in harmonic space (around 30 per cent by $\ell = 1024$); for completeness, we also apply a hard cut at $\ell = 1024$.

The maps are converted from the shear fields to convergence fields using the Kaiser–Squires reconstruction, described by equation (9), where both E and B-mode convergence maps are retained.

6.2 Power spectra, peaks, and neural compression

Power spectra: The power spectrum $C(\ell)$ of a field on the celestial sphere is defined via

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C(\ell)\delta_{mm'}\delta_{\ell\ell'}\,. \quad (15)$$

Here, $a_{\ell m}$ are the spherical harmonic coefficients of the field, δ is the Kronecker delta, and the expectation $\langle\rangle$ is with respect to random

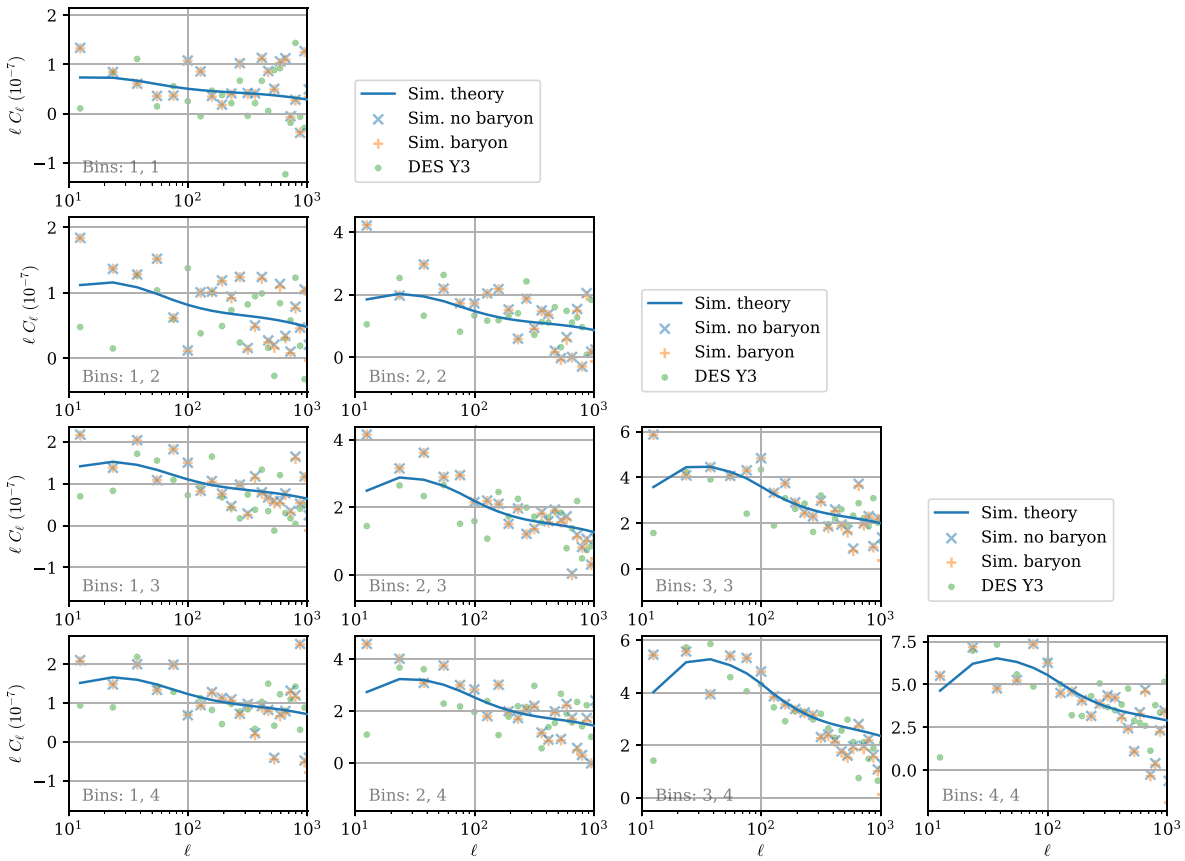

Figure 6. Power spectra C_ℓ (and cross-spectra between tomographic bins) of the baryonified simulation, the simulation without baryons, and the DES Y3 data. The theoretical power spectra, calculated with the same cosmological parameters as the two simulations, is shown for reference.

realizations. An unbiased estimate of this power spectrum is

$$\hat{C}(\ell) = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 \ . \tag{16}$$

It is the power spectrum of the shear field (not the convergence field) that we measure. We decompose the shear field into E- and B-modes (curl-free and divergence-free components, respectively), yielding shear power spectra $C_\ell^{\rm EE}$, $C_\ell^{\rm EB}$, and $C_\ell^{\rm BB}$. As with previous DES power spectra analysis, we use a pseudo-C_ℓ estimator that corrects for the effect of the sky mask. See Doux et al. (2022) for details. This correction is not actually necessary to give unbiased results, as the correction (or lack thereof) would be applied equally to the simulated and observed data. Following the pseudo-C_ℓ correction, we obtain $C_\ell^{\rm EE}$ and $C_\ell^{\rm BB}$ to use as our observed data vectors.

Fig. 6 shows the measured $C_\ell^{\rm EE} - C_\ell^{\rm BB}$ spectra for all tomographic bins, along with simulated spectra (discussed in Section 2.4). **Peaks:** A *peak* is a map pixel whose value exceeds that of its neighbouring pixels (typically there are eight such neighbours). For a given convergence map we create a histogram of the values of the convergence field at the peak pixels. Our histograms use 14 equally spaced bins and the range covered by these bins is chosen in advance so that each bin has at least ten peaks at a fiducial cosmology. We repeat this procedure on smoothed versions of our maps. This smoothing uses a top-hat filter; recall that in harmonic space the effect of such smoothing is to multiply the harmonic coefficients by

$$W_\ell(\theta_0) = \frac{P_{\ell-1}(\cos(\theta_0)) - P_{\ell+1}(\cos(\theta_0))}{(2\ell+1)(1-\cos(\theta_0))}, \tag{17}$$

where P_ℓ is the Legendre polynomial of order ℓ, θ_0 the smoothing angle, and ℓ the multipole. We consider eight smoothing angles θ_0 equally (logarithmically) spaced from 8.2 to 221 arcmin. We count the peaks for the convergence maps from each of the four tomographic maps of the DES Y3 weak lensing sample. In addition, we account for cross-correlation between bins by following Zürcher et al. (2022) and introducing 'cross-maps' $\kappa^{ij}(\theta, \phi)$. These are new maps obtained by combining two original convergence maps for different tomographic bins i and j (with $i > j$):

$$\kappa^{ij}(\theta,\phi) = \sum_{\ell=0}^{\ell_{\rm max}} \sum_{m=-\ell}^{\ell} \hat{\kappa}^i_{\ell m} \hat{\kappa}^j_{\ell m} Y_{\ell m}(\theta,\phi), \tag{18}$$

We compute the peak function in each of the resulting six new cross-maps (e.g. Fig. 7). Finally, before compressing the peak function, we adjust the peak counts of the noisy maps by subtracting the peak counts from a noise-only version of the maps.

Compression: The power spectra compression uses an ensemble of 12 multilayer perceptron (MLP) networks. As discussed in Section 2.3, we use an MSE loss function. All final fully connected layer outputs use the sigmoid activation function; as a result, compressed statistics are confined to a sensible domain. This choice of activation function therefore requires a rescaling of our parameters (so that their prior ranges lie well within the bounds of the sigmoid activation).

Each MLP has an input size of 560 ($= 10 \times 28 \times 2$), corresponding to the ten cross-correlations of the four tomographic bins, in 28 multipole (ℓ) bins, over the two components (EE and BB) of shear maps. The MLP network has ten hidden layers, each with

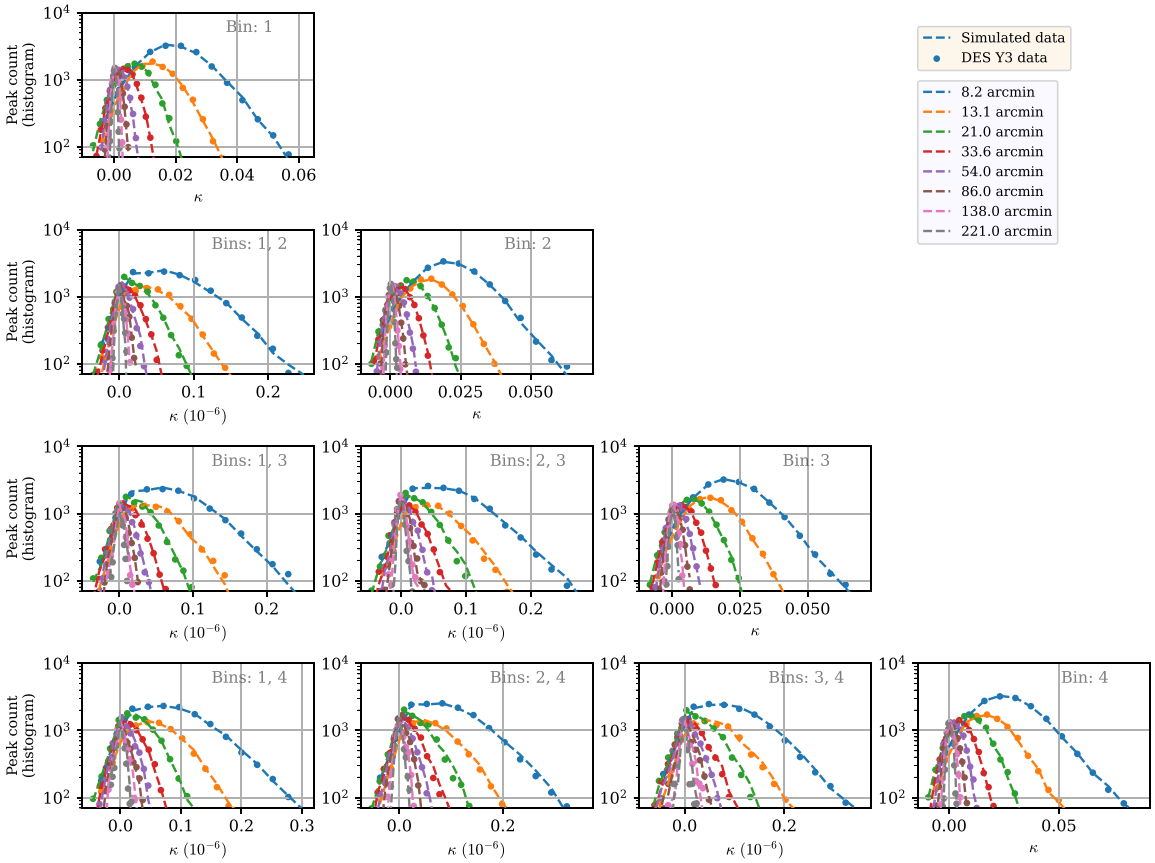

Figure 7. Peak count histograms from the DES Y3 data (solid circular markers) for each tomographic bin (rightmost plot on each row) and for the cross-maps. Each plot shows the observed peak counts for eight smoothing scales of the lensing map κ. For reference, we also show (dashed lines) histograms for a simulation that was chosen for its similarity with the actual observed data; it has $\Omega_{\rm m} = 0.29$, $S_8 = 0.82$, $w = -0.83$, and randomly sampled nuisance parameter values. The cross-maps have small κ values because equation (18) is not normalized.

256 nodes, with an embedded layer normalization and an ReLU (rectified linear unit) activation function at each layer output. The last layer reduces the output size to a single node corresponding to the selected parameter being compressed. Similarly to the CNN ensemble, there is a final sigmoid activation function on the output of the final layer. This MLP network is trained 12 times, each using the same data but different random network parameter initializations, and the resulting 12 predictions are averaged to yield an ensemble prediction. (We trained 12 times because this was clearly superior to training just once, and because it was convenient for the computer hardware being used; we do not claim optimality.)

The training input data is augmented using additive random Gaussian noise as a regularization measure. The noise added to each input bin ℓ_i is sampled from $\mathcal{N}(\mu_i, \sigma_i \times 10^{-3})$, where μ_i and σ_i are the mean and standard deviation of the values in each multipole bin ℓ_i observed across the training data set. Each network is optimized using the stochastic gradient decent Adam's optimizer with an MSE as the loss metric. The learning rate was initially set at 1×10^{-4} and decays exponentially with a decay rate of 0.1 per training step. The training is performed for up to 200 epochs with an early stopping criterion; this is described in more detail in 6.3.

Compression of the peak counts is done similarly; the only difference is that the MLP has an input size of 1120 ($= 10 \times 14 \times 8$).

6.3 Map-level (CNN) compression

This approach aims to infer cosmology directly from the map data. Here we implement a convolutional neural network (CNN) as a higher order statistic, using deep learning to compress relevant features directly from pixels; CNNs can be optimized to compress these features to a lower dimension in an informative way. We make no claim that our map-level compression is optimal (in the sense that the resulting parameter constraints are the best possible), but it is practical and does lead to significantly improved results.

Planar CNNs take flat two-dimensional images as input whereas our data (via the HEALPIX pixelization) are embedded on the sphere. There exist neural networks adapted to the geometry of the sphere (e.g. Defferrard et al. 2020; Ocampo, Price & McEwen 2022). For practicality, however, we have decided instead to perform separate analyses of several nearly flat rectangular patches on the sky. By using patches we lose large-angle correlation information, but this can be mitigated by combining (via concatenation of compressed data vectors) the CNN output with the compressed power spectrum output (as described in Section 6) – on large-angle scales we expect the signal to be near-Gaussian (and hence for these scales the power spectra are already maximally informative). We will refer to this combination as $C_\ell \times$CNN.

For our CNN approach we use patches from the sphere, flattened to two-dimensional images of 512×512 pixels. The 'nested' format

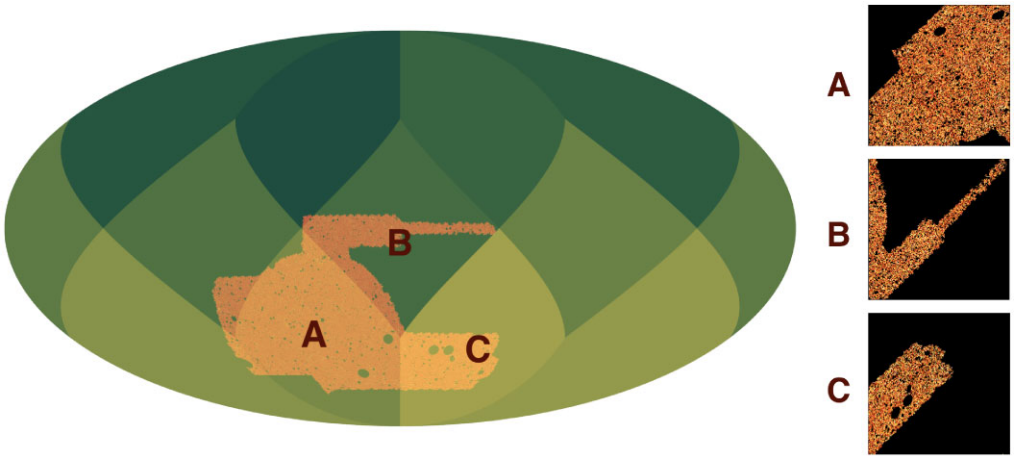

Figure 8. A demonstration of the CNN patching scheme for map-level compression, showing an example mock convergence (κ) HEALPIX map (in orange) with NSIDE = 512. The pixels of the convergence map are split into patches A, B, C based on the respective NSIDE = 1 pixel (in green). This produces the square patches seen to the right of the figure to be that are used to train the CNN ensemble. This patching scheme is lossless: there is a one-to-one match between the HEALPIX map pixels and the patch pixels.

for HEALPIX pixel ordering offers a natural method for extracting new patches; we form a patch by taking all the NSIDE = 512 resolution pixels that lie within a single superpixel defined by the minimum HEALPIX resolution NSIDE = 1, as shown in Fig. 8. This projection distorts the spherical geometry, which for traditional parameter estimation approaches would bias the inference. However, in our simulation-based method this transformation will also be applied consistently to the DES Y3 data, and therefore projection distortion will not bias any parameter estimation. Nevertheless, projection distortion makes the compression potentially suboptimal. Our chosen patch size is a compromise between loss of large-scale information (from having small patches) and projection distortion (from having large patches). The chosen scheme leads to the DES footprint being split into three patches (labelled A, B, C; a small subsection of the footprint is discarded).

To complement this patching scheme, we construct an ensemble of weak learning CNNs. Each patch has four dedicated networks, trained on the same data but with different network parameter initialization. Compressing each patch individually has the advantage of allowing each network to become familiar with the footprint of each patch. The resulting compression will be the weighted average of all 12 CNN networks (four per patch for three patches) for a single chosen parameter (with weights given by the sky fraction of each patch – see Fig. 8). This ensemble approach of averaging many simple CNNs has been shown to be a robust way to train networks that generalize well with smaller data sets to avoid overfitting; such an approach is advantageous considering the computational expense of constructing mocks.

The network is fed eight channels (E- and B-modes of the convergence maps for each of four tomographic bins); each channel supplies a patch of 512 × 512 pixels. The input maps are augmented during training by the same procedure described previously in 6.2, where the means and standard deviations are calculated for each individual pixel. The same setup for loss metric and optimisation function are applied identically to the networks as in 6.2.

Each of the eight convolutional layers in the network has 32 filters with kernel size 3 × 3. Each layer contains a batch normalization layer to stabilize training, an ReLU activation function, and an average pooling layer with a pooling size of 2 × 2 to downsample the input maps. The output of the convolutional blocks is then processed by a series of dense fully connected layers. These dense layers have sizes 20, 20, 10, 10, 5, and 5 (with ReLU activation). The final dense layer produces a single scalar output with a sigmoid activation function to produce a bounded network output.

The ensemble CNN was trained on 9264 DES Y3 mock data sets (three independent noise realizations of our 3088 original simulated convergence maps; see Section 5.4). Each individual CNN network in the ensemble was trained for up to 200 epochs, with early stopping to prevent overfitting. The stopping criterion is based on the MSE loss of a set of 3088 data with a different noise realization, which acts as a validation data set. We also use this validation data set when performing the neural density estimation task and this dual use of the validation data set has the potential to introduce bias. We have ruled out this possibility by checking that our inferred posteriors do not shift when using an additional different data set (to which the CNN is entirely blind during training) in place of the validation data when performing the neural density estimation task. This procedure attempts to mitigate any overfitting in two different ways.

The ensemble CNN trains in just under one hour per parameter on 12 Nvidia A100 GPUs using the NERSC Perlmutter cluster.

As a final step in the algorithm, the CNN output and the compressed power spectrum data vector are concatenated.

7 RESULTS

7.1 Prior probabilities

Table 1 summarizes the priors used in our cosmological inference.

The first (top) group includes the three target parameters. In our mock data, these parameters are *not* distributed according to our chosen prior. The parameters S_8 and Ω_m were sampled with active learning and have a particularly strange distribution in Fig. 4. The prior, therefore, must always be set explicitly when combining with the learned likelihood; furthermore, we must always include these parameters in our learned likelihood (unlike the parameters with implicit priors described below).

The middle group of parameters in Table 1 have implicit priors, i.e. the Gower Street simulations have these parameter distributions (see Section 4 for caveats). Even if these parameters are not explicitly included in the learned likelihood, they will be implicitly

Table 1. Prior and hierarchical probability distributions.

Parameter	Prior probability distribution
Ω_m	$\mathcal{U}(0.15, 0.52)$
S_8	$\mathcal{U}(0.5, 1.0)$
w	$\mathcal{U}(-1, \frac{1}{3})$
n_s	$\mathcal{N}(0.9649, 0.0063)$
h	$\mathcal{N}(0.7022, 0.0245)$
$\Omega_\mathrm{b}h^2$	$\mathcal{N}(0.022\,37, 0.000\,15)$
$\log(m_\nu)$	$\mathcal{U}[\log(0.06), \log(0.14)]$
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	$\mathcal{N}(-0.0369, 0.0076)$
$\bar{n}_i(z)$	$p_\mathrm{HYPERRANK}(\bar{n}_i(z)\|x_\mathrm{phot})$

marginalized during inference and the uncertainty in these parameters will be propagated to the final constraints on other parameters (see Section 2.5).

The final (bottom) group in Table 1 are the nuisance parameters, which are varied according to these prior distributions during mock shear map generation. Again, the uncertainty in these parameters is propagated through forward modelling in this simulation-based inference framework.

7.2 Simulation validation with Gower Street sims

7.2.1 Inference from mean mock data

We test that we can recover the correct 'input' parameter values from averaged data. This test is an analogue of the standard 'noise-free' inference, which is typically performed to demonstrate recovery of the input parameters. Instead, we take the average data vector from a set of mock simulations, perform inference, and validate that we recover the average parameter values from the same set of mock simulations.

The data vector used for this inference is the per-element mean of all compressed mock data vectors: $\bar{t}_j = \frac{1}{N}\sum_{i=0}^{N-1}(t_j)_i$ where the i index denotes individual mock data vectors (over the full parameter space for the Gower Street sims) and j indexes the elements of the data vector. This mean data vector uses all mock compressed data vectors that were used for the density estimation.

Fig. 9 shows the result of this test for the combination of mock power spectra C_ℓ and the map-level (CNN) compression. The mean parameter values are clearly recovered. We can confirm this test is also passed for the other combinations of data.

7.2.2 Coverage test results

As introduced and described in Section 2.4.1, coverage tests repeat the parameter inference procedure to test that the estimated posterior describes the correct probability for the parameters. We repeat the inference procedure, each time excluding one mock data vector from the neural likelihood estimation step. The likelihood is then evaluated using that held-out data vector, with the posterior then being evaluated and compared to the true parameter value.

We use the TARP package (Lemos et al. 2023) to estimate the coverage probabilities in the three-dimensional parameter space $\{\Omega_\mathrm{m}, S_8, w\}$ (rather than on the marginal posteriors individually). This code implements the 'Tests of Accuracy with Random Points'

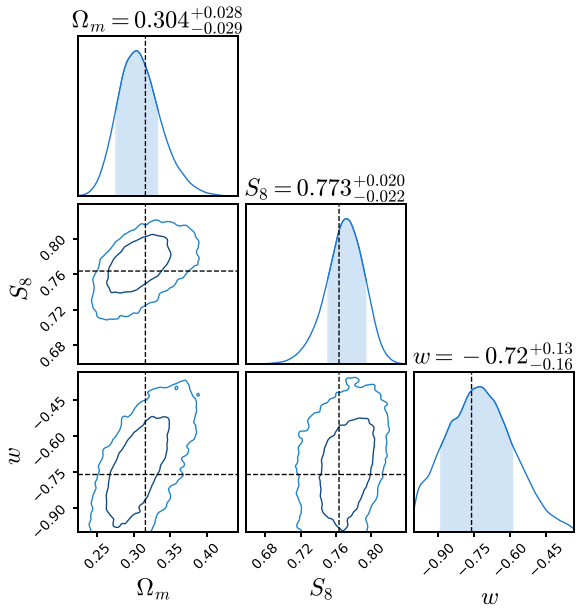

Figure 9. Inference with the mean mock data (combining mock power spectra and map-level CNN compressed data). The values for S_8 and Ω_m denoted by dashed lines are the average of the true parameter values from the same mock data. This is analogous to using noise-free data for inference when using a Gaussian likelihood.

(TARP) algorithm, which estimates coverage probabilities of generative posterior estimators. With this test, we used different, but still wide, priors on our target parameters ($S_8 \sim \mathcal{N}(0.78, 0.15)$, $\Omega_\mathrm{m} \sim \mathcal{N}(0.32, 0.07)$, $w \sim \mathcal{N}(-1, \frac{1}{3})$ truncated at the original prior limits); chosen to more closely match our simulated distribution (Fig. 4), so we get a high density of samples for this test. We repeat the neural likelihood estimation technique 175 times, in which we draw 6×10^3 samples from the learned posterior conditioned on a held-out data vector, and perform coverage testing using these Markov chain Monte Carlo (MCMC) samples as an input to TARP (v 0.1.1 using the 'manhattan' metric).

Fig. 10 shows the result of this procedure applied to the map-level CNN patch data (plots for the other observables show similar results). The expected coverage does indeed match the credibility level. This validates our neural likelihood estimation, showing that the posterior distribution (i.e. parameter uncertainties) truly represent the probabilities that the Universe has some true parameter value.

7.3 Robustness to mismodelling & residual systematic errors

7.3.1 Systematic error injection

We describe tests each of which confirms that the variation of some source of systematic error does not affect our results. In the language of machine learning statistics, these are robustness tests for (a specific type of) *distributional shift* – a mismatch between the training data and the deployment data.

The tests use the `CosmoGridV1` simulations suite (Kacprzak et al. 2023). We chose a set of one hundred simulations at the fiducial cosmology $\sigma_8 = 0.84$, $\Omega_\mathrm{m} = 0.26$, $w = -1$, $H_0 = 67.36$, $\Omega_\mathrm{b} = 0.0493$, $n_\mathrm{s} = 0.9649$. The `CosmoGridV1` simulations, like the Gower Street simulations, were created using the PKDGRAV3 code (Potter et al. 2017); they were created independently, however, and hence can serve as a further test that the correct input cosmology is recovered.

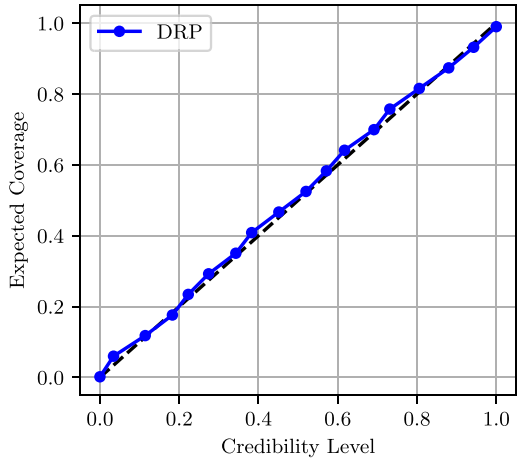

Figure 10. Coverage test result (using TARP) to validate the inference pipeline. Using repeated mock data parameter inference, the fraction of true values in the appropriate credible intervals matches the expected fraction. The figure shows the result for the map-level CNN patch compression; similar coverage tests were successful for the other observables. DRP is the 'Distance to Random Point' (see Lemos et al. 2023).

For each source of systematic error we generate two sets of mock data with different levels of systematic error included. We then apply our inference pipeline to each of these mock data sets and compare the resulting posterior probability distributions of the cosmological parameters.

To test for any overall shifts in the resulting posterior distributions, we use the average compressed data as the input to the neural likelihood estimation. For example, for the fiducial result, we measure each summary statistic for each of our selected mock `CosmoGridV1` data sets, then compress each separately, then average the resulting compressed statistic. This averaging mitigates the intrinsic variability that is expected between different data realizations.

Two sources of systematic error are tested: baryon feedback and source clustering bias variation.

(i) *Baryonic feedback* – Feedback effects can lead to suppression of structure on small cosmological scales. The Gower Street sims are N-body (dark matter only) simulations and do not include any baryonic astrophysics. In line with the standard DES weak lensing analyses, we cut scales that we think are likely to be affected by baryons and then test for the effect of possible baryonic contamination (e.g. Amon et al. 2021; Secco et al. 2022; Zürcher et al. 2022).
Hydrodynamical simulations are unfortunately too computationally expensive to generate a sufficient quantity of realistic mock data that include baryons. We therefore use the `CosmoGridV1` maps, as these include a baryon correction model; this model (Kacprzak et al. 2023) changes the density fields (in a post-processing step) to emulate baryon feedback.
The effect of baryons on the simulated power spectra can be seen in Fig. 6, which shows the measured power spectra from data in the fiducial set ('Sim. no baryon') and from data with baryon feedback included ('Sim. baryon'). The suppression can be seen at small scales (high ℓ).

(ii) *Source clustering bias variation* – Although we expect the effect to be small, it is possible that a different value for galaxy bias of the source galaxies could change our results. This is due to source galaxy clustering, which is known to change the predicted

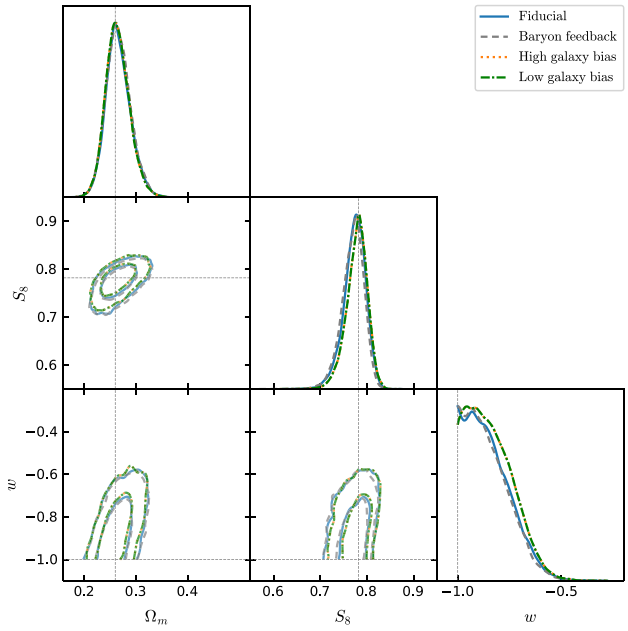

Figure 11. The marginal posterior distributions with independent `CosmoGridV1` simulated data with two sources of systematic error in the mock data: baryon feedback and varying source galaxy bias. Both of these are found to induce changes in the posterior below 0.3 σ in the marginal S_8-Ω_m plane (the standard Dark Energy Survey test). This test also shows that our pipeline recovers the true parameter values (straight dashed lines) with independent mock data. The shifts in the mean of marginal posterior distribution of w are all below 0.3σ.

observations (Section 5.4). We use a fixed value of galaxy bias $b = 1$ in our forward model, and hence we need to test that our results are not sensitive to a different true value in the observed data.
We generated two sets of simulated mock data, in addition to our fiducial mock data: one with a high-galaxy bias ($b = 1.5$) and one with a low-galaxy bias ($b = 0.5$).

Test results are shown in Fig. 11, which plots the posterior distribution for C_ℓ×CNN (power spectrum combined with map-level compression). Each of the two effects tested has a relatively small impact on the posterior. For the change in the S_8-Ω_m marginal posterior we use the standard DES criterion: we measure the shift in marginal posterior distribution relative to the standard deviation, finding that each of these systematic effects induces a shift below 0.3σ (the maximum level set by DES analyses). This test is also passed for C_ℓ (power spectra alone) and for C_ℓ×Peaks (power spectrum combined with peak counts).

Note that the true value of the input data $w = -1$ is on the boundary of the prior. We test the shift of the mean of the marginal posterior for w, finding that both systematic effects induce a shift of less than 0.3σ.

7.3.2 *Further systematic errors in lensing maps*

A potential source of contamination in the observed data is the misestimation of the point spread function (PSF). Failures in PSF modelling can cause errors in the measured shapes of galaxies, characterized by an additional ellipticity component $\delta\epsilon^{\rm sys}_{\rm PSF}$.

Jarvis et al. (2016) and Gatti et al. (2021) provide a model to describe $\delta\epsilon^{\rm sys}_{\rm PSF}$ that can be calibrated using *reserved stars*, i.e. those stars not used to train the original PSF model. We therefore could,

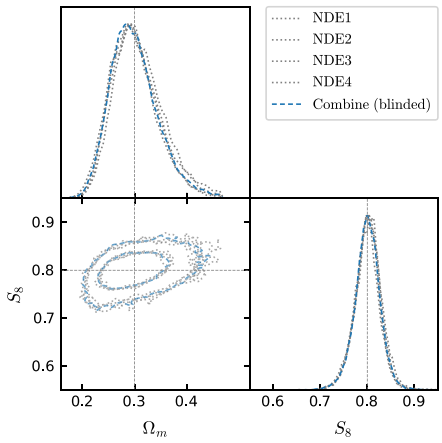

Figure 12. Neural density estimator ensemble convergence test using power spectrum C_ℓ data. Although using the real observed DES Y3 data, this was a blind test, which was achieved by shifting the posterior mean to a fiducial value ($\Omega_\mathrm{m} = 0.3$ and $S_8 = 0.8$).

following the procedure described in Gatti et al. (2024a), generate a map of $\delta\epsilon^{\mathrm{sys}}_{\mathrm{PSF}}$ per tomographic bin to be added to the fiducial shear maps; this could serve as a high, but in principle possible, contamination due to PSF errors. However, if we inject this PSF contamination at map level, we are overwhelmed by shot noise from our finite sample of reserved stars and this particularly affects small scales. We therefore do not use this approach (while we await further work that could accurately forward model PSF errors into our mock lensing maps).

We instead rely on alternative tests. In Gatti et al. 2021, the DES Y3 shear catalogue tests showed no evidence of additive biases due to PSF mismodelling. Furthermore, tests of reconstructed mass maps in Jeffrey et al. (2021b) showed no evidence of PSF residual errors.

7.4 Blinded data likelihood ensemble validation

To test the convergence of the neural density estimation (i.e. the likelihood learned from simulated data), we compare the different density estimates that comprise the ensemble. As described in Section 2.4.2, an insufficient number of simulated data realizations typically leads to significant differences between the neural likelihood estimates. With more simulations, the predictions from the ensemble converge.

Unlike the previous test on simulations, this test can be applied to the estimated likelihood evaluated for the actual observed data. This test was therefore done blind, as described in Section 7.5, and we confirmed that the test was passed before the full unblinded results were seen (Section 7.7).

Fig. 12 shows the posterior distributions from each likelihood in the ensemble for the observed DES data. In this example the data are the power spectra C_ℓ. This test was performed (and passed) before unblinding any of our results on data (including peaks and CNN map-level inference).

7.5 Blinding strategy

We used a blinding strategy (described below) to reduce the impact of confirmation bias. Blinding has been used by many DES analyses (Muir et al. 2020), but note that the approach of this paper made necessary some deviations from the standard DES blinding strategy.

(i) Some simulations used input cosmological parameters obtained from 'active learning' (see Section 4.2), and this required estimating the posterior distribution of the cosmological parameters using the simulations available at that point. These estimations were held within the computer code and were not revealed to the experimenters.

(ii) All training of the neural networks for compression and for density estimation was finalized without evaluation on any real observed data. The entire pipeline was run using a simulation (as if it were real data); results were checked for reasonableness and the neural network parameters were then frozen.

(iii) The uncompressed statistics from real data (the measured power spectra and peak counts) were checked for reasonableness.

(iv) The compressed statistics from real data were confirmed to be well within the convex hull of the scatterplot of compressed statistics obtained from the simulations (see Appendix A for a discussion of goodness-of-fit).

(v) The posterior distribution of cosmological parameters was inferred from observed data; this posterior was then shifted (by an amount that was kept within the computer code and was not available to the experimenters) to have a fiducial mean value. This shifted posterior was used in the likelihood ensemble validation of Section 7.4. It was also used to confirm that the posterior distribution had a figure of merit similar to that derived in a similar way from simulations (note that the figure of merit is sensitive to the width of the posterior but not to its mean).

(vi) Finally the shift to the posterior mean was removed, revealing the unblinded posterior.

7.6 Intrinsic alignments

7.6.1 Discussion

Our S_8 compression is not optimal. We find that including an additional compression of the map to informative summaries of A_{IA} (the intrinsic alignment amplitude) improves our posterior constraints on S_8. This shows that the original S_8 compression was missing information; this was to some extent expected (see Section 6.3 for a discussion).

This does not mean the intrinsic alignment nuisance parameters are incorrectly marginalized when using only the suboptimal S_8 summary, t_{S_8}. That is, the following desired property still holds:

$$p(t_{S_8}|S_8) = \int p(t_{S_8}|S_8, A_{\mathrm{IA}})\, p(A_{\mathrm{IA}})\, \mathrm{d}A_{\mathrm{IA}} \quad , \tag{19}$$

where $p(t_{S_8}|S_8)$ is a learned likelihood for S_8. What *is* true is that the posterior $p(S_8|t_{S_8}, t_{A_{\mathrm{IA}}})$ is tighter than $p(S_8|t_{S_8})$; this is because t_{S_8} is a suboptimal summary statistic.

Despite the possibility of improved cosmological constraints, we nevertheless do not include in our main analysis the compressed A_{IA} summary statistic, $t_{A_{\mathrm{IA}}}$. The primary reason for this choice is that including this statistic results in a posterior distribution so tight that the NDE ensemble test fails; the density of simulations in that region of parameter space becomes too low.

Even if this test had not failed, there would be further reasons to not include this additional information. The intrinsic alignment NLA model has been tested with direct two-point correlation measurements only down to scales of $\sim 5\,h^{-1}$Mpc (e.g. Johnston et al. 2019; Singh et al. 2023); at smaller scales linear galaxy bias modelling is insufficient. At the peak of our redshift distribution of source galaxies, around $z \sim 0.6$, our angular scale cuts correspond to a physical scale of $\sim 3\,h^{-1}$Mpc. We may have some confidence that the NLA model continues to hold at such small scales (unless there are unexpected higher-order contributions); nevertheless, if the

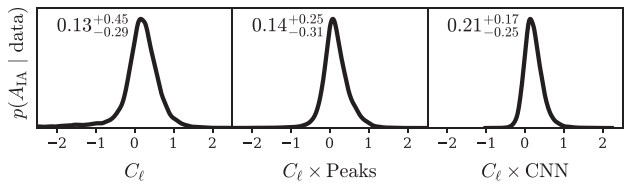

Figure 13. Marginal posterior distribution of the amplitude of intrinsic alignment $A_{\rm IA}$ for the three DES Y3 combinations: power spectra C_ℓ, peaks with power spectra $C_\ell\times$Peaks, and map-level inference $C_\ell\times$CNN. These constraints use additional data and a different w prior compared to our main cosmological results (see Section 7.6 for discussion), to show that the inferred intrinsic alignments are reasonable from this analysis.

constraints on S_8 are strongly affected by $t_{A_{\rm IA}}$ then it is prudent to exclude this additional information.

7.6.2 *Results*

Although we do not include the $t_{A_{\rm IA}}$ compressed statistics in our analysis for cosmological constraints (Section 7.7), here we present the marginal posteriors for $A_{\rm IA}$ using the $t_{A_{\rm IA}}$ statistics as a 'sanity check', i.e. to confirm that the inferred $A_{\rm IA}$ values are reasonable.

Fig. 13 shows the marginal posteriors for the intrinsic alignment amplitude $A_{\rm IA}$ for each of our standard data combinations: power spectra, peaks with power spectra, and map-level inference. To reduce the NDE dimension, we implicitly marginalize w, so that the w prior is given by $\mathcal{N}(-1, 1/3)$ for values of $-1 < w < -1/3$ (see Section 2.5 for discussion). This change does not particularly impact intrinsic alignment inference, and, furthermore, the aim of this inference is just to confirm that the $A_{\rm IA}$ values are reasonable.

The results for $A_{\rm IA}$ are all consistent, with a slight preference for low positive values, but still consistent with $A_{\rm IA} = 0$. This result is also consistent with the results from existing DES two-point analyses (e.g. Amon et al. 2021; Doux et al. 2022; Secco et al. 2022).

7.7 DES Y3: cosmological constraints

We present results using the three data combinations previously described: power spectra (C_ℓ), peaks and power spectra ($C_\ell\times$Peaks), and map-level inference ($C_\ell\times$CNN).

As described in Section 2.3, these data (summary statistics) x are compressed to lower-dimensional summary statistics $t = \mathcal{F}(x)$. In each case the compression function $\mathcal{F}$ is a neural network (or ensemble of networks), optimized for the given summary statistic.

We target the parameters $\Omega_{\rm m}$, $S_8(\equiv \sigma_8(\Omega_{\rm m}/0.3)^{1/2})$, and w, and thus the compressed data for a given summary statistic has three elements. When we combine the data (e.g. $C_\ell\times$Peaks) we concatenate the compressed data vectors (e.g. $t = {\rm concat}[t_{C_\ell}, t_{\rm Peaks}]$), giving a compressed data vector with six elements. All other parameters, including cosmological and nuisance parameters, are implicitly (and correctly) marginalized; see Section 2.5 for details.

Fig. 14 shows the marginal two-dimensional posterior distribution for the three data combinations. Credible intervals derived from the one-dimensional marginals were calculated using the GetDist package (Lewis 2019) and are listed in Table 2. This figure and table show the main result of this paper.

All of these results use a full simulation-based (likelihood-free) inference pipeline to infer cosmological parameters.

We can also compare our most constraining result, that from $C_\ell\times$CNN (map-level inference combined with power spectrum),

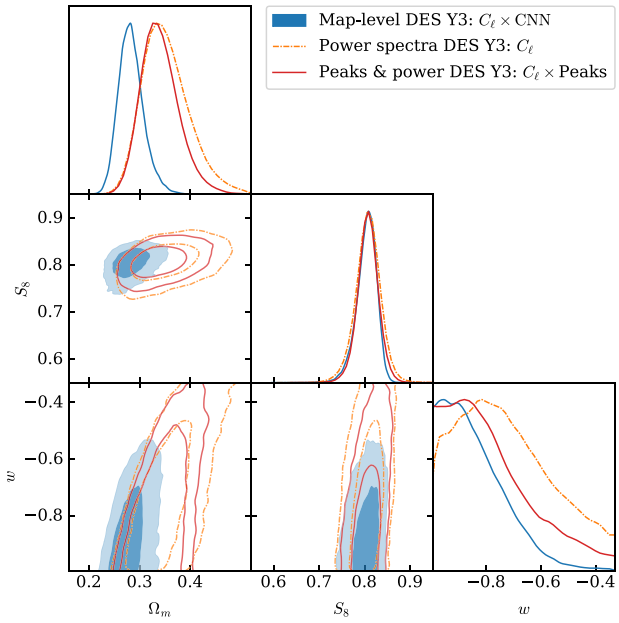

Figure 14. Posterior probability distribution for $\{\Omega_{\rm m}, S_8, w\}$ obtained from simulation-based inference using three DES Y3 data combinations: power spectra C_ℓ, peaks with power spectra $C_\ell\times$Peaks, and map-level inference $C_\ell\times$CNN.

Table 2. Comparison of summary statistics used in this analysis (power spectrum, peaks, and map-level inference): 68 per cent credible intervals from the marginal posterior probability distributions of $\Omega_{\rm m}$, S_8, and w.

	Power spectrum (this work): C_ℓ	Peaks & power (this work): $C_\ell\times$Peaks	Map-level DES Y3 (this work): $C_\ell\times$CNN
$\Omega_{\rm m}$	$0.352^{+0.035}_{-0.053}$	$0.340^{+0.030}_{-0.042}$	$0.283^{+0.020}_{-0.027}$
S_8	$0.807^{+0.027}_{-0.025}$	0.807 ± 0.023	$0.804^{+0.025}_{-0.017}$
w	< -0.661	< -0.740	< -0.803

to existing data and likelihoods. We compare to the Planck cosmic microwave background (CMB) data; here we use the Planck likelihood code with model priors amended to match our analysis choices (except for $\Omega_{\rm b}$, as our prior here was motivated by Planck). We also compare to the DES Y3 likelihood for real-space weak lensing two-point correlation functions; here also we have matched the analysis choices and priors (where appropriate).

Fig. 15 compares our analysis with these two alternative cosmological inference pipelines. We find our results to be consistent with these existing data and analysis pipelines (i.e. the Planck and the DES Y3 weak lensing likelihoods). We recover values lower than Planck not only for S_8 (such a tension between CMB and weak-lensing results is already well-known, e.g. Amon & Efstathiou 2022) but also for $\Omega_{\rm m}$.

Table 3 presents credible intervals derived from the one-dimensional marginals in Fig. 15.

The Planck re-analysis uses the 2018 TTTEEE-lowE likelihood Planck Collaboration (2021) with settings matching those used in Abbott et al. (2023). All priors, except for $\Omega_{\rm b}$, were matched to our analysis (Table 1). As our choice of $\Omega_{\rm b}$ prior was informed by Planck, for the Planck re-analysis we use a prior matching (Abbott et al. 2023).

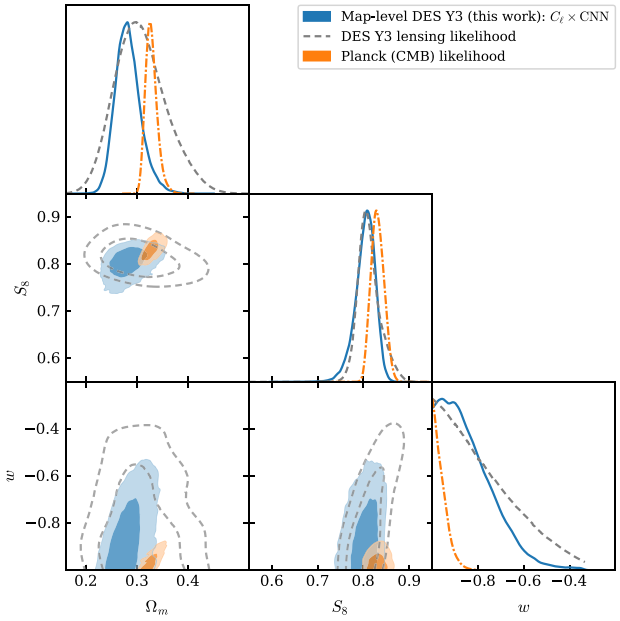

Figure 15. Comparison of the $C_\ell \times$CNN result (map-level compression) with results both from the Planck CMB likelihood and from the standard DES weak gravitational lensing (two-point correlation function) likelihood (both of which have been subject to reanalysis to match prior choices).

Table 3. Comparison with existing analyses: 68 per cent credible intervals from the marginal posterior probability distributions of Ω_m, S_8, and w. We compare the $C_\ell \times$CNN result (map-level compression) with the results from both the standard DES weak gravitational lensing (2-point correlation function) likelihood and the Planck CMB data. The standard DES likelihood and Planck likelihood results have used prior choices matched to our analysis to allow comparison.

	Map-level DES Y3 (this work): $C_\ell \times$CNN	DES Y3 lensing likelihood*	Planck (CMB) likelihood*
Ω_m	$0.283^{+0.020}_{-0.027}$	$0.303^{+0.040}_{-0.051}$	$0.328^{+0.009}_{-0.013}$
S_8	$0.804^{+0.025}_{-0.017}$	$0.813^{+0.020}_{-0.029}$	$0.831^{+0.014}_{-0.015}$
w	< -0.803	< -0.707	< -0.954
			* reanalysed

We do not include shear ratio information (e.g. Sánchez et al. 2021) for the DES Year 3 weak lensing re-analysis.

We sample the posterior using the Planck and the DES Y3 weak lensing likelihoods with PolyChord (Handley, Hobson & Lasenby 2015). For the parameters h, Ω_b, n_s, and Ω_m, we use a flat prior during the MCMC sampling and then importance reweight to the desired prior as a post-processing step.

8 CONCLUSION

We have presented the DES Y3 simulation-based inference results, in which we have used power spectra, peak counts, and map-level compression/inference to constrain parameters of the wCDM model.

Our approach seeks to improve both accuracy and precision.

For improved accuracy, we use simulation-based inference as this allows us to forward model realistic effects in our simulated data. For those effects about which there is uncertainty (measurement biases, photometric redshift uncertainties, effects of neutrinos, and intrinsic alignments of galaxies), we randomly vary the effect in our mock data according to our prior probability. This is relatively straightforward in this inference framework; for example, the marginalization over possible redshift distributions $n(z)$ amounts to the marginalization of approximately one thousand nuisance parameters.

We have tested that our results are robust to certain types of model misspecification (namely source galaxy biasing and baryon feedback). We have also tested that our recovered posterior distributions have the correct coverage; this is made possible by our fast (almost amortized) inference pipeline.

For improved precision, we include weak lensing statistics beyond standard two-point statistics. In particular, we directly compress the weak lensing mass map (i.e. *dark matter map*, Kaiser & Squires 1993), and then use simulation-based inference to construct a likelihood for the compressed map. Combining this compressed mass map and the compressed power spectra yields improved constraints on the parameters of the wCDM model (compared to our results using compressed power spectra alone). Table 2 lists the 68 per cent credible intervals of the marginal posteriors per parameter.

These improvements are often quoted in terms of the Figure of Merit, given by FoM $= (\det\Sigma)^{-1/2}$ for posterior covariance Σ; this is a measure of inverse volume (i.e. tightness) of the posterior probability. For the weak lensing parameter combination $\{S_8, \Omega_m\}$ we improve the FoM by a factor of 2.26, while for the dark energy parameter combination $\{\Omega_{DE}, w\}$ we improve the FoM by a factor of 2.48. (In this latter parameter combination we included neutrinos in Ω_m and neglected the small photon radiation contribution, so $\Omega_{DE} = 1 - \Omega_m$ for a flat Universe.)

The challenge for future analyses is to improve the modelling accuracy so that we can either apply less conservative data cuts or spend effort improving the map level compression. This will likely come from more realistic simulations. Further improvements in the forward model would allow less conservative use of the available data, further improving the precision beyond even these new results.

The improvements in precision bring an increased responsibility to maintain accuracy. The principal challenge presented by this approach is achieving sufficiently realistic data modelling, which, if accomplished, will substantially increase the potential for discovery.

ACKNOWLEDGEMENTS

We thank F. Lanusse and B. Wandelt for helpful comments and discussions at many points during this project.

NJ was supported by STFC Consolidated Grant ST/V000780/1 and by the Simons Collaboration on Learning the Universe. The Gower Street simulations were generated under the DiRAC project p153 'Likelihood-free inference with the Dark Energy Survey' (ACSP255/ACSC1) using DiRAC (STFC) HPC facilities (www.dirac.ac.uk).

JP was supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures programme. JA was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programmes (Grant agreement no. 101018897 CosmicExplorer).

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award HEP-ERCAP-0027266.

Funding for the DES Projects was 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 AstroParticle 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 Zürich (ETH), 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, NFS'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 was 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. IF was partially funded by the CERCA programme 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.

DATA AVAILABILITY

The Gower Street simulations are available at www.star.ucl.ac.uk/GowerStreetSims/; the metacalibration lensing catalogue is available at https://des.ncsa.illinois.edu.

The MCMC samples from the parameter posteriors (i.e. chains) will be made available upon publication of the accepted paper here: www.star.ucl.ac.uk/GowerStreetSims/.

REFERENCES

Abbott T. M. C. et al., 2018, ApJS, 239, 18
Abbott T. M. C. et al., 2023, Phys. Rev. D, 107, 083504
Alsing J., Wandelt B., 2018, MNRAS, 476, L60
Alsing J., Heavens A., Jaffe A. H., 2017, MNRAS, 466, 3272
Alsing J., Wandelt B. D., Feeney S. M., 2018a, preprint (arXiv:1808.06040)
Alsing J., Wandelt B., Feeney S., 2018b, MNRAS, 477, 2874
Alsing J., Charnock T., Feeney S., Wand elt B., 2019, MNRAS, 488, 4440
Amon A., Efstathiou G., 2022, MNRAS, 516, 5355
Amon A. et al., 2021, Phys. Rev. D, 105, 023514
Angulo R. E., Hahn O., 2022, Living Rev. Comput. Astrophys., 8, 1
Asgari M. et al., 2021, A&A, 645, A104
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bishop C., 1994, Mixture Density Networks. Aston University, United Kingdom
Blazek J. A., MacCrann N., Troxel M. A., Fang X., 2019, Phys. Rev. D, 100, 103506
Bridle S., King L., 2007, New J. Phys., 9, 444
Castro P. G., Heavens A. F., Kitching T. D., 2005, Phys. Rev. D, 72, 023516
Charnock T., Lavaux G., Wandelt B. D., 2018, Phys. Rev. D, 97, 083004
Cordero J. P. et al., 2022, MNRAS, 511, 2170
Cranmer K., Pavez J., Louppe G., 2015, preprint (arXiv:1506.02169)
Defferrard M., Milani M., Gusset F., Perraudin N., 2020, preprint (arXiv:2012.15000)
Doux C. et al., 2022, MNRAS, 515, 1942
Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241
Flaugher B. et al., 2015, AJ, 150, 150
Fluri J., Kacprzak T., Refregier A., Amara A., Lucchi A., Hofmann T., 2018, Phys. Rev. D, 98, 123518
Fluri J., Kacprzak T., Lucchi A., Refregier A., Amara A., Hofmann T., Schneider A., 2019, Phys. Rev. D, 100, 063514
Fluri J., Kacprzak T., Lucchi A., Schneider A., Refregier A., Hofmann T., 2022, Phys. Rev. D, 105, 083518
Gatti M. et al., 2021, MNRAS, 504, 4312
Gatti M. et al., 2022, MNRAS, 510, 1223
Gatti M. et al., 2024a, Phys. Rev. D, 109, 063534
Gatti M. et al., 2024b, MNRAS, 527, L115
Goodfellow I. J., Bengio Y., Courville A., 2016, Deep Learning. MIT Press, Cambridge, MA
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 453, 4384
Heavens A. F., Jimenez R., Lahav O., 2000, MNRAS, 317, 965
Hermans J., Delaunoy A., Rozet F., Wehenkel A., Begy V., Louppe G., 2021, preprint (arXiv:2110.06581)
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Huff E., Mandelbaum R., 2017, preprint(arXiv:1702.02600)
Jarvis M. et al., 2016, MNRAS, 460, 2245
Jeffrey N., Wandelt B. D., 2020, Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS). University College London, available at: https://discovery.ucl.ac.uk/id/eprint/10142290/
Jeffrey N., Alsing J., Lanusse F., 2021a, MNRAS, 501, 954
Jeffrey N. et al., 2021b, MNRAS, 505, 4626
Jimenez Rezende D., Mohamed S., 2015, preprint (arXiv:1505.05770)
Johnston H. et al., 2019, A&A, 624, A30
Kacprzak T., Fluri J., Schneider A., Refregier A., Stadel J., 2023, J. Cosmol. Astropart. Phys., 2023, 050
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kingma D. P., Salimans T., Jozefowicz R., Chen X., Sutskever I., Welling M., 2016, in Advances in Neural Information Processing Systems. p. 4743
Kingma D. P., Salimans T., Poole B., Ho J., 2024, in Proc. 35th International Conference on Neural Information Processing Systems (NIPS'21). Curran Associates Inc., Red Hook, NY
Knabenhans M. et al., 2021, MNRAS, 505, 2840
Kullback S., Leibler R. A., 1951, Ann. Math. Stat., 22, 79

Lemos P., Coogan A., Hezaveh Y., Perreault-Levasseur L., 2023, 40th International Conference on Machine Learning, Vol. 202. p. 19256
Lewis A., 2019, preprint (arXiv:1910.13970)
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Li X. et al., 2023a, Phys. Rev. D, 108, 123518
Li S.-S. et al., 2023b, A&A, 679, A133
MacCrann N. et al., 2022, MNRAS, 509, 3371
Morganson E. et al., 2018, PASP, 130, 074501
Muir J. et al., 2020, MNRAS, 494, 4454
Myles J. et al., 2021, MNRAS, 505, 4249
Ocampo J., Price M. A., McEwen J. D., 2022, preprint (arXiv:2209.13603)
Papamakarios G., Pavlakou T., Murray I., 2017, Adv. Neural Inf. Process. Syst., 30
Papamakarios G., Sterratt D., Murray I., 2019, in Chaudhuri K., Sugiyama M., eds, Proc. Machine Learning Research Vol. 89, Proc. Twenty-Second International Conference on Artificial Intelligence and Statistics. PMLR, p. 837, available at: https://proceedings.mlr.press/v89/papamakarios19a.html
Peel A., Lalande F., Starck J.-L., Pettorino V., Merten J., Giocoli C., Meneghetti M., Baldi M., 2019, Phys. Rev. D, 100, 023508
Planck Collaboration VI, 2021, A&A, 652, C4
Potter D., Stadel J., Teyssier R., 2017, Comput. Astrophys. Cosmol., 4, 2
Prangle D., Blum M. G., Popovic G., Sisson S., 2014, Aust. NZ J. Stat., 56, 309
Ribli D., Pataki B. Á., Zorrilla Matilla J. M., Hsu D., Haiman Z., Csabai I., 2019, MNRAS, 490, 1843
Riess A. G. et al., 2022, ApJ, 934, L7
Sánchez C. et al., 2021, Phys. Rev. D, 105, 083529
Secco L. F. et al., 2022, Phys. Rev. D, 105, 023515
Sellentin E., Heavens A. F., 2018, MNRAS, 473, 2355
Sellentin E., Heymans C., Harnois-Déraps J., 2018, MNRAS, 477, 4879
Sellentin E., Loureiro A., Whiteway L., Lafaurie J. S., Balan S. T., Olamaie M., Jaffe A. H., Heavens A. F., 2023, Open J. Astrophys., 6, 31
Sevilla-Noarbe I. et al., 2021, ApJS, 254, 24
Sevilla I. et al., 2011, in Meeting of the APS Division of Particles and Fields (DPF 2011)
Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24
Singh S., Shakir A., Jagvaral Y., Mandelbaum R., 2023, MNRAS, 530, 3515
Taylor P. L., Kitching T. D., Alsing J., Wandelt B. D., Feeney S. M., McEwen J. D., 2019, Phys. Rev. D, 100, 023519
Tram T., Brandbyge J., Dakin J., Hannestad S., 2019, J. Cosmol. Astropart. Phys., 2019, 022
Zürcher D. et al., 2022, MNRAS, 511, 2075

APPENDIX A: DISCUSSION: GOODNESS-OF-FIT IN COMPRESSED DATA SPACE

As part of the unblinding procedure, we validated that the compressed data from the observations $\mathbf{t}_O$ were in-distribution $q(\mathbf{t})$ where q can be characterized by the simulated data samples.

Put another way, we showed that $\mathbf{t}_O$ fell within the distribution of simulated $\mathbf{t}_i$. Furthermore, we found that the probability mass Q inside the isoprobability hypersurface defined by $q = q(\mathbf{t}_O)$ was not extreme. A low value of $p = 1 - Q$ would have implied a poor goodness-of-fit for this test. In this sense, we have a high goodness-of-fit in compressed data space.

We only discuss this type of test in an appendix, as success in this goodness-of-fit test may be misleading. Passing this test (in compressed data space) is 'necessary but not sufficient' and hence this test comes with a warning: in general, having $\mathbf{t}_O$ within the distribution $q(\mathbf{t})$ does not require $\mathbf{x}_O$ to be within the corresponding distribution $q(\mathbf{x})$. For $\mathbf{t} = F(\mathbf{x})$ (where F reduces the dimensionality of the data i.e. is a compressor), the compressed data space may not present any discrepancies even if the uncompressed observed data is truly out of distribution.

In our work, we have mitigated any out-of-distribution errors through a series of validation tests that give us high confidence in our results. Nevertheless for simulation-based inference it is useful to study methods for testing goodness-of-fit in high-dimensional spaces; this is related to ongoing work on anomaly detection in such spaces.

[1]*Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK*
[2]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[3]*Oskar Klein Centre for Cosmoparticle Physics, Stockholm University, Stockholm SE-106 91, Sweden*
[4]*Faculty of Physics and Astronomy, Astronomical Institute, Ruhr University Bochum, German Centre for Cosmological Lensing, D-44780 Bochum, Germany*
[5]*Nordita, KTH Royal Institute of Technology and Stockholm University, Hannes Alfvéns väg 12, SE-10691 Stockholm, Sweden*
[6]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA*
[7]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[8]*Université Grenoble Alpes, CNRS, LPSC-IN2P3, F-38000 Grenoble, France*
[9]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[10]*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA*
[11]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[12]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[13]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[14]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue Madison, WI 53706-1390, USA*
[15]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15312, USA*
[16]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[17]*Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ-20921-400, Brazil*
[18]*Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain*
[19]*Department of Physics, Duke University Durham, NC 27708, USA*
[20]*NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA*
[21]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[22]*Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA*
[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, D-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]*Kavli Institute for Particle Astrophysics & Cosmology, P.O. Box 2450, Stanford University, Stanford, CA 94305, USA*
[29]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[30]*Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, SP-13083-859, Campinas, Brazil*

[31]*Department of Physics, University of Genova and INFN, Via Dodecaneso 33, I-16146 Genova, Italy*
[32]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[33]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[34]*Brookhaven National Laboratory, Bldg 510, Upton, NY 11973, USA*
[35]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
[36]*Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, F-31400 Toulouse, France*
[37]*Excellence Cluster Origins, Boltzmannstr. 2, D-85748 Garching, Germany*
[38]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*
[39]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*
[40]*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*
[41]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[42]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[43]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[44]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[45]*Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway*
[46]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[47]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[48]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[49]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[50]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA*
[51]*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*
[52]*Center for Astrophysics|Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[53]*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*
[54]*LPSC Grenoble - 53, Avenue des Martyrs, F-38026 Grenoble, France*
[55]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[56]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ-20921-400, Brazil*
[57]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[58]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.