Map-level baryonification: unified treatment of weak lensing two-point and higher-order statistics

Zhou, A.J.; Gatti, M.; Anbajagane, D.; Dodelson, S.; Schaller, M.; Schaye, J.

Map-level baryonification: unified treatment of weak lensing two-point and higher-order statistics

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Received: *May 14, 2025*
Accepted: *August 20, 2025*
Published: *September 24, 2025*

Map-level baryonification: unified treatment of weak lensing two-point and higher-order statistics

Alan Junzhe Zhou,[a,c] **Marco Gatti**,[b,c] **Dhayaa Anbajagane**,[b,c] **Scott Dodelson**,[b,c,d] **Matthieu Schaller**[e,f] **and Joop Schaye**[e]

[a] *Department of Physics, University of Chicago, Chicago, IL 60637, U.S.A.*
[b] *Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, U.S.A.*
[c] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, U.S.A.*
[d] *Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, U.S.A.*
[e] *Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands*
[f] *Lorentz Institute for Theoretical Physics, Leiden University, P.O. Box 9506, NL-2300 RA Leiden, The Netherlands*

E-mail: ajzhou@uchicago.edu, mgatti@uchicago.edu, dhayaa@uchicago.edu, dodelson@fnal.gov, schaller@strw.leidenuniv.nl, schaye@strw.leidenuniv.nl

Abstract: Precision cosmology benefits from extracting maximal information from cosmic structures, motivating the use of higher-order statistics (HOS) at small spatial scales. However, predicting how baryonic processes modify matter statistics at these scales has been challenging. The baryonic correction model (BCM) addresses this by modifying dark-matter-only simulations to mimic baryonic effects, providing a flexible, simulation-based framework for predicting both two-point and HOS. We show that a 3-parameter version of the BCM can jointly fit weak lensing maps' two-point statistics, wavelet phase harmonics coefficients, scattering coefficients, and the third and fourth moments to within 2% accuracy across all scales $\ell < 2000$ and tomographic bins for a DES-Y3-like redshift distribution ($z \lesssim 2$), using the FLAMINGO simulations. These results demonstrate the viability of BCM-assisted, simulation-based weak lensing inference of two-point and HOS, paving the way for robust cosmological constraints that fully exploit non-Gaussian information on small spatial scales.

Keywords: cosmological simulations, semi-analytic modeling, weak gravitational lensing, power spectrum

ArXiv ePrint: 2505.07949

https://doi.org/10.1088/1475-7516/2025/09/073

JCAP09(2025)073

Contents

1 Introduction

Weak gravitational lensing directly probes the total matter distribution in the Universe, making it a powerful cosmological tool. Unlike galaxy clustering, which is biased by galaxy-matter relationships, weak lensing is sensitive to all matter, both dark and baryonic. Measurements of

JCAP09(2025)073

two-point weak lensing statistics have provided some of the tightest constraints on cosmological parameters, particularly $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$, which captures the degeneracy between the amplitude of matter fluctuations (σ_8) and matter density (Ω_m) [1–7].

There is growing interest in extracting non-Gaussian features from the lensing map, both at the field level [8–13] and via higher-order statistics (HOS). In particular, many HOS have been successfully applied to data have shown significant gains in constraining power (e.g., [14–20]). They are also expected to improve our constraining power on extended cosmological models, such as those with primordial non-Gaussianities [21] or modified theories of gravity [22, 23]. Unlike two-point statistics, however, most HOS lack analytical predictions and rely on realistic weak lensing simulations for modeling.

Weak lensing observables — both standard two-point and HOS — probe low redshifts ($z < 1$) and small physical scales ($r < 10\,\mathrm{Mpc}$), making them particularly sensitive to non-linear structure formation and astrophysical processes such as supernova and active galactic nucleus feedback, collectively referred to as baryonic feedback. These astrophysical processes alter both the distribution and thermodynamics of baryons, and affect the dark matter distribution through gravitational coupling (e.g., [24–33]). These effects are important on small and intermediate scales, extending into the quasi-linear regime ($\lesssim 5\,\mathrm{Mpc}$), where they contribute significantly to the lensing signal. If unaccounted for, baryonic effects can bias cosmological inferences [15, 34–37]. For example, the observed discrepancy in S_8 between weak lensing and CMB measurements could arise from such mis-modeling rather than a fundamental deviation from the standard cosmological model [e.g., 36]. While a very conservative approach is to remove scales affected by baryons or to implement projection schemes that mitigate their impact [1, 2, 38, 39], accurate modeling remains the preferred solution.

On the modeling side, the baryonic imprints on the matter density field are typically studied using high-resolution cosmological hydrodynamical simulations [see 40, for a review]. However, there is significant freedom in defining the "subgrid" physics in these simulations — processes like star formation and collisional cooling that occur below the simulation's resolution scale. Different but *equally well-motivated* subgrid prescriptions result in a wide range of predictions for the distribution of galaxy properties [e.g., 28, 29, 41–49], and consequently for the baryonic imprints on the non-linear matter power spectrum [e.g., 26, 33, 36, 50, 51]. While it is computationally unfeasible to explore broad cosmological and baryonic parameter spaces with these simulations,[1] they are nonetheless essential for validating and calibrating simpler, more practical parametric models of baryonic effects.

Several parametric models have been developed to flexibly describe the impact of baryons on weak lensing observables. A well-known example is HMCode2020 [53], an augmented halo model designed to accurately predict the non-linear matter power spectrum across a wide range of cosmologies. To account for baryonic feedback, HMCode2020 modifies the concentration-mass relation through a parametric model that accounts for mass and redshift-dependence and is fitted to the matter power spectra of the BAHAMAS simulations [54]. More recent

[1] Simulation suites like CAMELS [52] provide a wide variety of simulated predictions spanning this broad parameter space. However, these simulations are performed in cosmologically small volumes ($V \sim (25\,\mathrm{Mpc}/h)^3$) and so cannot be reliably used to explore the impact of baryons on statistics summarizing the large-scale matter distribution.

JCAP09(2025)073

examples include the power spectrum emulator trained on the FLAMINGO simulations [32] and a wavenumber-dependent rescaling of the difference between the non-linear and linear gravity-only power spectra [33, 36]. Another successful approach is the baryonic correction model (BCM), a phenomenological method that mimics baryonic effects by perturbing the density field of dark matter-only (DMO) simulations via post-processing [55–57]. These modified outputs can then be used to build emulators of the matter power spectrum that include such baryonic imprints. The BCM has demonstrated sufficient flexibility to reproduce a wide range of feedback scenarios seen in hydrodynamical simulations [56, 58, 59], and has already been applied to analyze two-point statistics from wide-field surveys [60–62].

Most of the recipes described above have been extensively tested in the context of two-point statistics. In principle, the BCM approach — via its use of N-body simulations — allows the construction of emulators for any observable, including HOS. However, the robustness of the model predictions for beyond two-point functions, in the context of analyzing current and future weak lensing datasets, remains largely unproven. Some studies have begun to explore this direction: [27] proposed a halo-based model which corrects the baryonic effects on the two and three-point functions. Ref. [63] showed that the BCM can simultaneously fit the matter power spectrum and bispectrum of hydrodynamical simulations for snapshots at different redshifts. Ref. [64] compared weak lensing peak counts from the full hydrodynamical simulation IllustrisTNG to those from a baryon-corrected version of its dark matter-only (DMO) counterpart, IllustrisTNG-Dark. They found that the BCM provided an insufficient fit when the noise levels match those expected from future surveys like LSST and Euclid. However, their analysis relied on BCM parameters calibrated solely to the power spectrum, whereas a joint fit that included peak counts could have improved performance. Moreover, they did not consider other types of HOS. More recently, [57] performed a joint fit of second-, third-, and fourth-moments of the matter density and gas pressure fields measured in IllustrisTNG, finding statistical agreement with the model on scales above $\sim 1\,\mathrm{Mpc}$ and across multiple redshifts. However, this test was also limited to projected slices at fixed redshifts and did not validate against realistic weak lensing observables.

We show that the BCM is sufficiently flexible for simulation-based analyses by applying it to the FLAMINGO simulations, a state-of-the-art suite that features multiple baryonic feedback models, matching DMO runs, and is calibrated against observed $z = 0$ galaxy stellar mass function and the galaxy cluster gas fractions [65, 66]. We baryonify the DMO simulations, generate realistic weak lensing maps, and build an emulator for both two-point and higher-order statistics (HOS) — including scattering transform coefficients [67], wavelet phase harmonics [68], and third- and fourth-order moments. The emulator accurately recovers the summary statistics of the hydrodynamical runs, and we show that appropriate BCM parameter choices can jointly fit all statistics across the FLAMINGO feedback models down to $\ell \approx 2000$.

The paper is organized as follows. Section 2 discusses the BCM model. Section 3 describes the FLAMINGO simulations and the DES Y3-like tomographic weak lensing convergence maps. Section 4 defines the relevant HOS and how they are measured in simulations. Section 5 describes the procedure for fitting the BCM model to the FLAMINGO feedback variants, and section 6 presents the results. We conclude in section 7. Supporting figures and tables are provided in the Appendix.

JCAP09(2025)073

2 Baryonification of simulations with BaryonForge

The BCM introduced in [55] is a phenomenological framework that does not require specifying any small-scale subgrid physics. The method is parameterized by pairs of halo density profiles: one profile representing a halo in a DMO simulation, and another representing the halo in a hydrodynamical simulation (DMB). The latter includes contributions from dark matter (DM) and baryons (B), in the form of stars, gas, satellite galaxies, and dark matter. There exist different models for the exact form and parameterization of these profile components [43, 55, 56, 58, 69]. Here we use the model of [59], which is a slightly modified version of the model in [56]. We refer the reader to those works for detailed discussions. Briefly, the total matter in the DMO case follows a Navarro-Frenk-White (NFW) profile [70, 71] with a power-law cut-off at large scales, $r > 4R_{200c}$ [72]. The stellar density follows an exponential profile, while the gas density follows a modified generalized NFW (gNFW) profile. The collisionless matter (dark matter and satellite galaxies) is modeled as a modified NFW profile that accounts for adiabatic relaxation due to the presence of the star and gas components. In our implementation, each halo is assigned a NFW concentration using the semi-analytic model of [73].

The DMO and DMB profiles can be expressed via the quantity $M(r)$, the mass enclosed within a radius r, which is a monotonically increasing function. The DMO and DMB enclosed mass profiles are then combined to compute a displacement function,

$$\Delta d(r) = M_{\rm DMB}^{-1}\left(M_{\rm DMO}(r)\right) - r. \tag{2.1}$$

where the functions above are implicitly also functions of the halo mass, M_{200c}. Here, the function $M_{\rm DMB}^{-1}$ takes a mass as its argument, and returns the radius that encloses that mass under the DMB model. This displacement function represents the radial offset required to transform a DMO matter distribution into a DMB one.

The BCM proceeds as follows. We start with a DMO simulation with a halo catalog. For each halo above a the mass threshold $10^{13}M_\odot/h$,[2] we identify all (dark matter) particles within a predetermined radius. Then, given the halo mass M_{200c} and the radial separation of the particles from the halo, we evaluate the displacement $\Delta d(r)$. Next, we radially[3] offset the particles by $\Delta d(r)$ from the halo center. The displacements are generally positive on the scale of the halo radius, as the dark matter particles must be moved outwards to mimic the impact of extended gas distributions, and are generally negative on scales a hundredth of the halo radius, where the dark matter particles must coalesce to mimic the centrally peaked stellar distribution. The final particle distribution represents a density field that mimics the impact of baryonic processes.

One can then vary the parameters controlling the mass profiles and thereby obtain different corrections to the simulated density field. The profiles specified by a given model can then also be used to predict other quantities, such as the integrated stellar/mass as a function

[2] We chose a threshold of $10^{13}\,M_\odot/h$ for computational reasons: the halo mass function is dominated by low-mass halos, which would exponentially increase the pipeline's runtime. The impact of excluding these small halos is discussed in the results section.

[3] [57] found the impact of anisotropic (elliptical) offsets on the moments of the density field to be sub-percent (see their figure 7).

JCAP09(2025)073

of halo mass. The BCM approach can simultaneously fit variations in the matter power spectrum alongside other quantities such as the scaling relations of gas and star fractions with halo mass, the halo gas profiles, etc. [56, 59, 74, 75].

Our discussions thus far describe the application of BCM to three-dimensional particle snapshots. However, for the purpose of weak lensing, it is computationally expensive to apply the BCM to each individual snapshot. To produce accurate lensing convergence maps, we require fine redshift resolution, e.g., $\mathcal{O}(100)$ snapshots, which increases the computational load of this task. The snapshot-based approach is also particularly expensive for the purpose of simulation-based inference, where of order $\mathcal{O}(10^4)$ simulations are needed to infer cosmological constraints. Therefore, we use the method outlined in [57] which circumvents the computational cost by applying the BCM directly on 2D density shells on the full-sky.[4] This is done by rewriting the model in terms of *projected* profiles and distances, rather than three-dimensional ones. Instead of particles, our units are the 2D pixels in the full-sky maps. For each pixel, we apply displacements from all halos — where the given pixel is contained with an aperture of $20R_{200c}$ around each selected halo — and accumulate the contributions from all halos to compute the total displacement of the pixel. This extension allows for efficient modeling of higher-order correlations in weak lensing observables by making the BCM more scalable, and therefore makes the method more easily applicable to large datasets from current and upcoming cosmological surveys.

All baryonification in this work is carried out using the publicly available BaryonForge[5] package [57]. We direct readers to [57] for more details on the model and method choices. For reference, generating 12 different baryonified simulations, each simulation consisting of 60 tomographic density field shells with $N_{\rm side} = 2048$, takes approximately 4 hours on an AMD EPYC 7763 CPU node with 128 cores and 512 GB of memory. However, we only use a fraction of the memory as the memory footprint of the procedure is set by the density shells and the halo catalogs. For this work, the total memory footprint is less than 10 GB across all shells.

The BCM parameters are listed in table 1, along with their ranges and descriptions. As we will see, the three most important parameters — the ones that can reproduce most of the variation across the hydrodynamic variants — are:

- M_c: a mass scale at which the characteristic slope of the gas profile is $\beta = 1.5$. Halos above/below this mass-scale have more/less cored gas profiles.

- $\theta_{\rm ej}$: a physical scale quantifying how extended the gas distribution is, expressed in units of R_{200c} of the halo. We will see that this ratio is of order 5, so the gas has quite an extended distribution.

- η: the power-law scaling of the stellar mass fraction with halo mass, $M_{\rm star}/M_{200c} \propto M_{200c}^{\eta}$.

The exact parameters we vary are primarily informed by the analysis of [59]. Compared to their work we also add the redshift scaling parameters ν_X (though, we will show below these

[4] An earlier version of this approach is also discussed in [76], but does not contain a number of modeling elements introduced in [57].

[5] https://github.com/DhayaaAnbajagane/BaryonForge.

JCAP09(2025)073

Param	Prior	Description
ν_{M_c}	[-5, 5]	The redshift scaling of the gas mass scale.
$\nu_{\theta_{ej}}$	[-5, 5]	The redshift scaling of the ejection radius.
γ	[0.1, 8]	The slope of the generalized NFW profile at $r \sim R_{ej}$.
$\log_{10}\theta_{co}$	[-4.0, -0.1]	The logarithm of the radial scale of the collapsed gas, $R_{ej} = \theta_{co} R_{200c}$.
$\log_{10}\eta_\delta$	[-3.0, 0.0]	The power-law scaling for the central galaxy stellar fraction-halo mass relation.
δ	[2, 20]	The slope of the generalized NFW profile at $r \gg R_{ej}$.
μ_β	[-3, 3]	The mass-dependent scaling of β, which controls the slope between $R_{co} < r < R_{ej}$.
$\log_{10}\frac{M_c}{M_\odot}$	[12.0, 16.0]	The logarithm of the mass-scale where the characteristic gas slope transitions.
$\log_{10}\eta$	[-3.0, 0.0]	The logarithm of the power-law scaling of stellar fraction-halo mass relation.
θ_{ej}	[0.8, 20]	The radial scale of the ejected gas, $R_{ej} = \theta_{ej} R_{200c}$.

Table 1. The input parameters to the profiles of the BCM. For each parameter, we list its prior range and a description of the parameter's physical meaning. See [57] for the specific equations that these parameters govern. The prior ranges are similar to those in [57, 69] and are used in section 5.1 to construct samples of the baryon-corrected maps and to train the BCM emulator.

are not required for accurate predictions) and η. The latter is added to give the model more flexibility in varying the relative fractions of the gas and star components in a halo.

3 Simulations

3.1 FLAMINGO simulations

Our analysis relies on the FLAMINGO suite of cosmological hydrodynamical simulations, all of which share the initial conditions of a DMO simulation. We provide a brief summary here; for more details, see [65, 66]. The simulations were run using the SPHENIX smoothed particle hydrodynamics implementation [77] within SWIFT [78]. FLAMINGO includes radiative cooling and heating [79], star formation [80], time-dependent stellar mass loss [81], massive neutrinos [82], and kinetic supernova and stellar feedback [83, 84]. The simulations are calibrated to match the galaxy stellar mass function at $z = 0$ and the gas fraction of low-z clusters and groups of clusters inferred from X-Ray and weak lensing observations [66]. Halos are identified using a three-dimensional friends-of-friends (FoF) algorithm and we utilize the corresponding FoF masses of these halos.[6]

[6]While it is more correct to use spherical overdensity (SO) masses for the baryonification method, we did not have the relevant SO masses for every simulation of interest during the start of this project when the baryonified maps were generated. We have thus used FoF masses instead, which were available for all simulations. While the use of FoF masses over SO masses will lead to some difference (and cause small shifts in the inferred parameters), it does not affect our ability to quantify the flexibility of the baryonification model for the statistics of interest.

JCAP09(2025)073

We focus on AGN feedback, implemented in two forms: thermal [85], where particles in the AGN vicinity are heated, and kinetic jet-like, where particles are given momentum kicks along the black hole (BH) spin axis [86]. We consider nine variations in AGN feedback strength from the models available in the FLAMINGO simulations, all in 1 Gpc^3 volumes:

- **Fiducial**: this model features thermal, AGN feedback without jets. Outflows follow paths of least resistance, forming buoyant high-entropy gas bubbles in clusters. Energy is injected into the nearest gas particle, reducing feedback at large distances compared to mass-weighted schemes. The AGN heating temperature, ΔT_{AGN}, is calibrated to match observations such as the cluster gas fractions. This model is labeled Fiducial (L1_m9) per FLAMINGO convention.

- Thermal AGN — cluster gas and galaxy stellar mass function variations: we include **6 variations** of the fiducial model where AGN feedback remains thermal, but the simulations feature different cluster gas fractions (Δf_{gas}) and/or galaxy stellar mass functions ($\Delta\Phi_*$), shifted by a specified number of observational standard deviations (σ). These variations lead to different AGN feedback strengths. The models are labeled fgas+2σ, fgas−2σ, fgas−4σ, fgas−8σ, M−σ, and M−σ_fgas−4σ.

- Jet-like AGN Feedback: the **two additional models**, Jet and Jet_fgas−4σ, replace the Fiducial thermal AGN feedback with kinetic, jet-like feedback. Compared to the Fiducial model, Jet has slightly weaker stellar feedback, while Jet_fgas−4σ features higher jet velocity.

We apply the baryonification procedure described in section 2 to the full-sky spherical overdensity shells of the DMO simulation to generate baryonified realizations. See appendix A of [65] for details on how the lightcone is generated. Our procedure for making lensing maps is detailed below in section 3.2. All simulations — both hydrodynamical and baryonified — adopt the best-fit ΛCDM cosmology from DES Y3 "3×2pt + All Ext." [87], which is $h = 0.681, \Omega_{\mathrm{m}} = 0.306, \sigma_8 = 0.807$. For the remaining parameters, see table 4 in [65].

Finally, to investigate the cosmology dependence of baryonic feedback in section 6, we also consider simulations run at the "LS8" cosmology [65, 88], which has a lower amplitude of the power spectrum: $S_8 = 0.766$ compared to 0.815 for the fiducial cosmology. The simulations include DMO (LS8_DMO), the fiducial AGN (LS8), and the strong AGN (LS8_fgas-8σ) variants. We did not apply the BCM to the LS8_DMO simulation and used these three simulations exclusively for the cosmology dependence analysis in section 6.

3.2 DES-like weak lensing convergence maps

We post-process the FLAMINGO simulations to generate weak lensing convergence maps, κ. Each FLAMINGO simulation provides lightcone outputs in the form of HEALPix maps [89] as a function of redshift. The native HEALPix resolution of $N_{\mathrm{side}} = 16384$ is degraded to $N_{\mathrm{side}} = 2048$ for this project (corresponding to a pixel scale of approximately 1.4 arcminutes), with shell thickness $\Delta z = 0.05$. We produce noiseless, full-sky convergence maps assuming the DES Y3 weak lensing redshift distributions in figure 1 from [90, 91]. Maps are made

JCAP09(2025)073

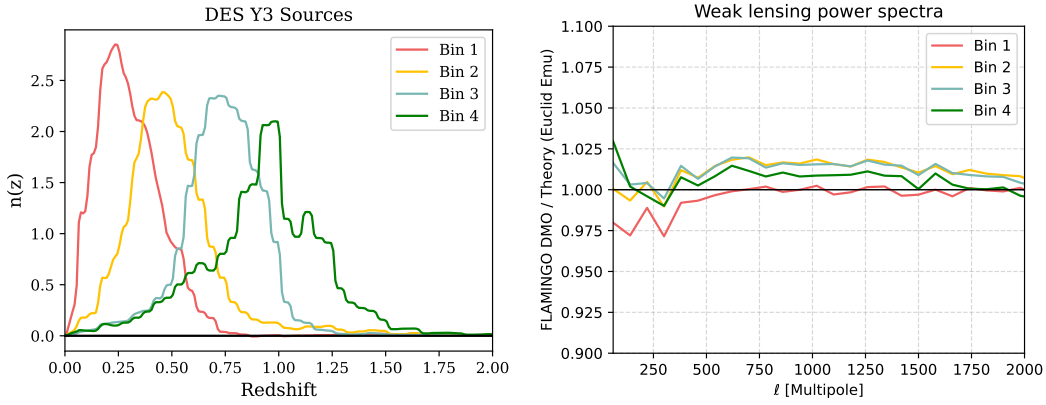

Figure 1. *Right:* redshift distribution of the four tomographic bins of the DES Y3 wak lensing sample considered here. *Left:* power spectra of the 4 noiseless, full-sky convergence maps produced from the FLAMINGO DMO simulation, using the DES Y3 redshift distributions, compared to predictions from EuclidEmu [94].

assuming the Born approximation.[7] Mimicking the DES Y3 tomographic sample, we generate 4 convergence maps by summing the shells weighted by the appropriate redshift distribution, enabling us to study baryonic effects as a function of redshift. Figure 1 shows the resulting 2D DMO power spectra in the four bins compared to theoretical predictions from the Euclid Emulator [94], with agreement at the 1–2% level, consistent with the emulator's accuracy. We do not add the signal from the intrinsic alignments (IA) of galaxies to our maps, and the coupling of IA and baryonic feedback is minimal [95, 96].

Since our focus is on the relative impact of baryonic feedback compared to the DMO simulation, we do not forward-model the DES Y3 noise or mask, and work exclusively with full-sky, noiseless convergence maps. In total, we use three sets of tomographic maps:

- $\kappa_{\mathbf{dmo}}$: one set from the DMO simulation,
- $\kappa_{\mathbf{flamingo}}$: 9 sets from the nine FLAMINGO feedback variants,
- $\kappa_{\mathbf{bary}}$: $\sim 5,000$ sets, generated by baryonifying $\kappa_{\mathbf{dmo}}$ following the BCM prescription in section 2, discussed further in section 5.1.

4 Statistics

We now describe the theory, measurement, and scale cuts of the statistics used here: the power spectra ($C(\ell)$), the scattering transform coefficients (ST), the wavelet phase harmonic coefficients (WPH), and the third- (m_3) and the fourth-moments (m_4) of the κ maps. As an example, the statistics of $\kappa_{\mathbf{flamingo}}$ are shown in figure 2 in colored lines. We also summarize

[7] [92] show that post-born corrections, estimated via a full ray-tracing algorithm, result in of $\mathcal{O}(10\%)$ corrections to statistics on small-scales. This is a statement on the absolute accuracy of a simulation compared to the observed lensing process. In this work we are focused on relative differences between simulations, where the relevance of this effect is greatly suppressed. Note also that [93, see their figure 1] show the Born approximation produces a consistent (within 1%) result for the angular power spectrum of the matter field for scales up to $\ell < 10^4$.

JCAP09(2025)073

all the statistics in table 2 at the end of the section. All statistics are computed on maps of $N_{\rm side} = 1024$, corresponding to a pixel scale of $3.44\prime$. At $z = 0.5$, which is where the lensing kernel peaks [e.g., 2], this corresponds to a comoving distance of $d = 1.9$Mpc. This is computed assuming the cosmology of the Flamingo simulation used in this work.

4.1 Power spectrum

The power spectrum of a κ map is defined as

$$C(\ell) = \frac{1}{2l+1} \sum_m |\hat{\kappa}_{\ell m}|^2 \,, \tag{4.1}$$

where $\hat{\kappa}_{\ell m}$ is the spherical harmonic transform of κ that is downgraded to $N_{\rm side} = 1024$. We consider the auto-power spectra of the 4 DES redshift bins, linearly binned in $\ell \in [20, 2020]$.

A main objective is to validate BCM's weak lensing statistics to spatial scales smaller than the previous DMO analyses. For example, the DES Y3 weak lensing simulation-based inference with HOS was limited to $\ell_{\rm max} = 1000$ [19, 97], because the analysis assumed a DMO model and could not model smaller scales contaminated by baryonic feedback. Here, we aim to validate the BCM to $\ell_{\rm max} = 2000$. We do not attempt to model $\ell > 2,000$ due to 1) increased systematic effects and a low signal-to-noise ratio in real data, and 2) the computational challenge of producing enough simulations with low shot noise at high ℓ in the context of simulation-based inference.

4.2 Scattering Transform Coefficients (ST)

Wavelets are localized oscillatory functions that act as band-pass filters in both real and Fourier space. Unlike Fourier analysis which is a global transformation, wavelets provide spatial localization while simultaneously capturing frequency information. A wavelet family consists of wavelets of different scales and orientations, and can be used to extract multi-scale patterns in images while preserving directional information [98].

Given a template wavelet $\psi(\boldsymbol{r})$,[8] we can generate new wavelets of different sizes and orientations via

$$\psi_{j,l}(\boldsymbol{r}) \propto \psi\left(\frac{1}{2^j} R^{-1}_{l\pi/L} \boldsymbol{r}\right), \quad j \in \{\mathbb{Z} \mid 0 \le j < J\}\,, \quad l \in \{\mathbb{Z} \mid 0 \le l < L\}\,. \tag{4.2}$$

Here j and l label the scale and the orientation of the derived wavelets. Each wavelet covers $\sim 2^j$ image pixels, and larger j corresponds to larger scales. The operator R^{-1} in the argument is the rotation matrix with the angle of rotation given by the subscript, with the wavelets covering L directions uniformly distributed over $[0, \pi)$.

The scattering transformed κ maps are constructed by iteratively applying the wavelet convolution and the modulus operator [67, 98]. The first and second-order scattering transformed κ maps are

$$\kappa^{{\rm ST},(1)}_{j,l} = |\kappa \star \psi_{j,l}| \,, \tag{4.3}$$

$$\kappa^{{\rm ST},(2)}_{j,l,j',l_\Delta} = \big|\, |\kappa \star \psi_{j,l}| \star \psi_{j',l-l_\Delta} \big| \,, \tag{4.4}$$

[8] For scattering transform, we use the Morlet wavelet under the convention of [67].

JCAP09(2025)073

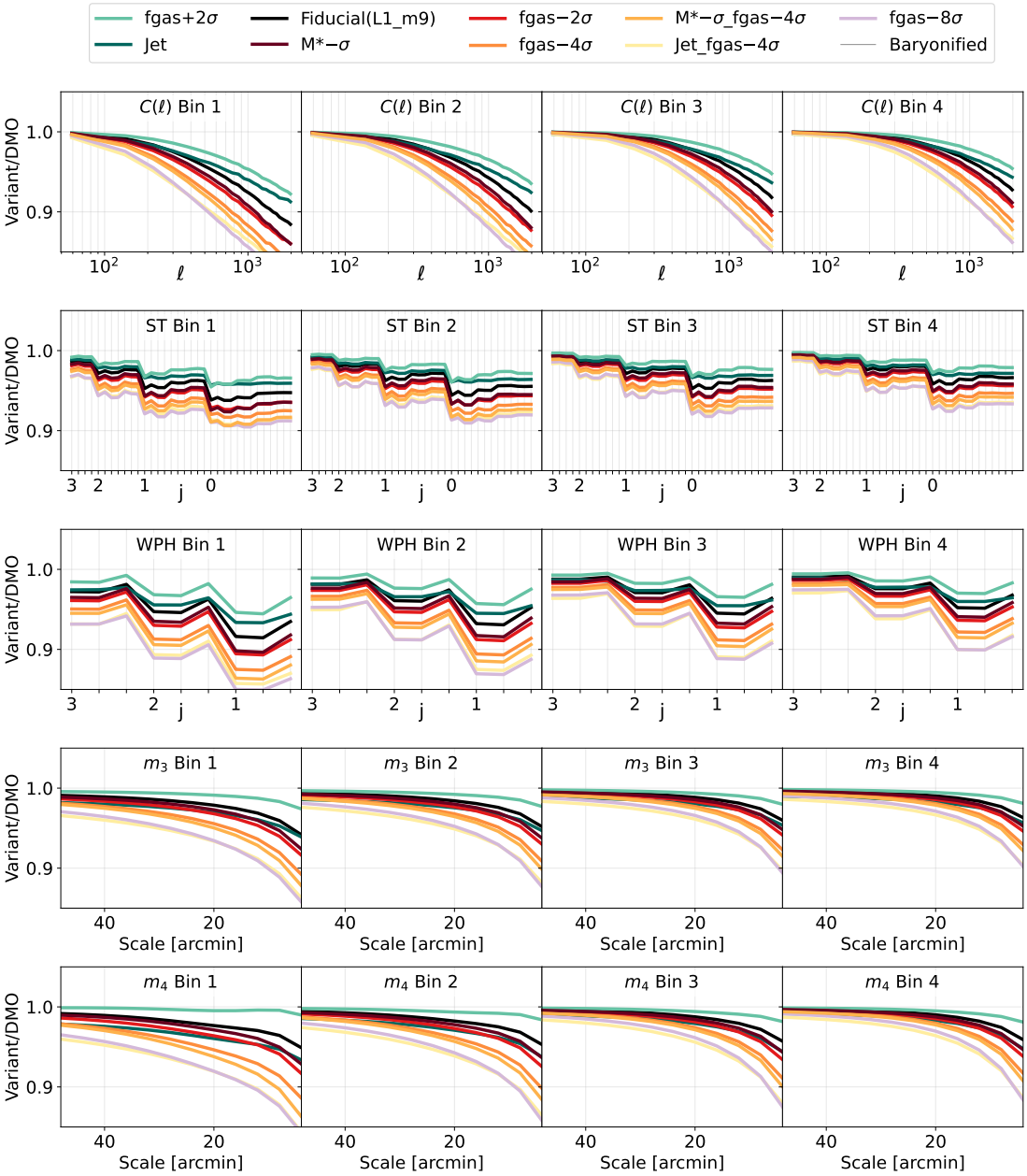

Figure 2. Summary statistics of the FLAMINGO κ maps (section 3). Rows correspond to different statistics, and columns to redshift bins (modeled after DES Y3 data). The y-axis shows the FLAMINGO statistics relative to the statistics of the DMO simulation of the same initial condition. On the x-axis, spatial scales decreases to the right. ST and WPH are ordered by the wavelet size (j) as a proxy for their characteristic scale. In FLAMINGO, baryon feedback physics suppresses all statistics on all scales, and is the strongest in low redshift bins.

JCAP09(2025)073

where $\star$ denotes convolution. The modulus preserves the first-order field statistics while introducing the necessary non-linearity to extract information beyond the mean without the instability to outliers that higher-order moments would cause. The higher-order convolution extracts information about spatial distributions $\kappa^{\mathrm{ST},(1)}$, and captures HOS by measuring the "clustering of clustered structures," with each order n potentially containing information up to 2^n-point functions [67, 99]. For the second-order field, we typically require $j \leq j'$ to ensure that the second wavelet extracts information at scales equal to or larger than the first. Since κ is a homogeneous and isotropic field, we compress κ^{ST} through spatial and angular averaging [67, 100]. More specifically, the ST coefficients are given by

$$s_j^{(1)} = \langle \kappa_{j,l}^{\mathrm{ST},(1)} \rangle_{p,l} \,, \tag{4.5}$$

$$s_{j,j',l_\Delta}^{(2)} = \langle \kappa_{j,l,j',l_\Delta}^{\mathrm{ST},(2)} \rangle_{p,l}, \tag{4.6}$$

where $\langle \cdot \rangle$ denotes the expectation value, and the subscripts indicate averaging over pixels (p) and orientations (l). These coefficients compactly and robustly summarize the information in the hierarchical scales of the clustering signal beyond the power spectrum.

To measure ST, we downgrade the κ maps to $N_{\mathrm{side}} = 1024$, and divide them into $N_{\mathrm{side}} = 8$ HEALPix pixels, each gnomonically projected to a 256×256 Cartesian patch with $3.44\prime$ resolution. Each patch is centered on an $N_{\mathrm{side}} = 8$ pixel with surrounding HEALPix pixels masked to prevent double counting. We measure ST on 1/8 of the sky using the `scattering_transform` package[9] with parameters $J = 4$ and $L = 3$, leading to 4 $s^{(1)}$'s and 30 $s^{(2)}$'s per redshift bin. The final ST is calculated independently for each tomographic bin and averaged across all Cartesian patches.

Since ST does not have a sharp scale cutoff like $C(\ell)$, we empirically determine the characteristic scale of each ST, and perform consistent scale cut in appendix A. All 34 ST coefficients per redshift bin are found to predominantly probe scales below $\ell_{\mathrm{max}} = 2000$.

4.3 Wavelet Phase Harmonics Coefficients (WPH)

WPH is another class of wavelet-based summary statistics. Again, we use $\psi_{j,l}(\boldsymbol{r})$ to denote wavelets of different sizes (j) and orientations (l), but with a different waveform — the bump-steerable wavelets [101]. WPH statistics are obtained by first convolving κ with $\psi_{j,l}(\boldsymbol{r})$, then taking its phase harmonics, and finally computing the moments of the resulting fields [68, 101]. Since κ is isotropic, we also average over the orientations of the wavelets. The equations for the first- and second-order WPH reduce to

$$S_j^{(1,1)} = \langle \mathrm{Cov}(\kappa \star \psi_{j,l},\, \kappa \star \psi_{j,l}) \rangle_l \,, \tag{4.7}$$

$$S_j^{(0,0)} = \langle \mathrm{Cov}(|\kappa \star \psi_{j,l}|,\, |\kappa \star \psi_{j,l}|) \rangle_l \,, \tag{4.8}$$

$$S_j^{(0,1)} = \langle \mathrm{Cov}(|\kappa \star \psi_{j,l}|,\, \kappa \star \psi_{j,l}) \rangle_l \,. \tag{4.9}$$

$S^{(0,0)}$ and $S^{(0,1)}$ contain both Gaussian and non-Gaussian information due the nonlinearity of the phase harmonics (in this case, the modulus) operation. $S^{(1,1)}$ contains only Gaussian information, albeit with a different smoothing scheme compared to the power spectrum.

[9] https://github.com/SihaoCheng/scattering_transform.

JCAP09(2025)073

The measurement of WPH is similar to that of ST, on gnomonically projected patches of $3.44\prime$ resolution. We again work with Cartesian patches that are projected from the κ maps. On each patch, we compute WPH using the `PyWPH` package.[10] We consider wavelets with $J = 4$ and $L = 3$, leading to 4 of each $S^{(1,1)}$, $S^{(0,0)}$, and $S^{(0,1)}$ per redshift bin. We then average the WPH from each patch across 1/8 of the sky to reduce cosmic variance.

The scale cut for WPH is empirically determined in appendix A and is consistent with the $\ell_{\rm max} = 2000$ scale cut for the power spectra. The three $j = 0$ WPH are removed from subsequent analyses.

4.4 Higher order moments

The third and fourth moments of the convergence field are defined as

$$m_3(\theta) = \left\langle \left(\kappa \star G(\theta)\right)^3 \right\rangle_p , \tag{4.10}$$

$$m_4(\theta) = \left\langle \left(\kappa \star G(\theta)\right)^4 \right\rangle_p , \tag{4.11}$$

where $G(\theta)$ is a Gaussian kernel of width θ and p denotes pixels. m_3 measures skewness of κ. It contains only non-Gaussian information and is strictly zero for a Gaussian random field. m_4 measures the kurtosis of κ and has both a Gaussian (derivable from the power spectrum) and a non-Gaussian contribution following Wick's theorem. We set the minimum scale at $4\prime$ and use $\theta \in [4\prime, 48\prime]$ in increments of $4\prime$. This is consistent with the scale cuts of other statistics since a $4\prime$ Gaussian-smoothed κ has negligible power beyond $\ell = 2000$.

The modeling and the real data measurement of m_3 and m_4 are more sensitive to noise than two-point statistics. This is because 1) by the central limit theorem, for a sample with size N, while $\langle\kappa\rangle$ has an error $\propto N^{-1/2}$, for $\langle\kappa^n\rangle$, the effective standard error scales more like $\alpha N^{-1/2}$ where α grows superlinearly with n, and 2) the tail of outliers of the distribution contributes disproportionally, amplifying the fluctuations. Perturbation theory provides a similar perspective, where empirically $\langle\kappa^n\rangle \propto \langle\kappa^2\rangle^{n-1}$ [102]. Therefore an x uncertainty in $\langle\kappa^2\rangle$ propagates to $(n-1)x$ in $\langle\kappa^n\rangle$.

4.5 The statistics of FLAMINGO and baryonified convergence maps

In total, we consider 87 (348) data points per redshift bin (and in total), as summarized in table 2. Throughout this work, we consider summaries of the auto-correlation of the four tomographic bins, but do not include their cross-correlations for the sake of reducing the size of the data vector. We still test the ability of the baryonification model to fit all four tomographic bins, and are therefore adequately probing the model's redshift evolution (as it pertains to weak lensing analyses). The statistics for $\kappa_{\rm flamingo}$, relative to those of $\kappa_{\rm dmo}$, are shown in figure 2. In all FLAMINGO feedback variants, baryons suppress all statistics at all scales. The suppression is stronger at smaller scales and lower redshifts, consistent with previous works [e.g., 26, 33, 36, 50, 103, 104]. Relative to the fiducial model, the fgas+2σ variant shows weaker baryonic signatures, whereas the fgas−2σ and Jet_fgas−4σ variants induce stronger suppression in the statistics.

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

JCAP09(2025)073

Statistics		Equation	N per bin	Notes
ST	$s_j^{(1)}$	Eq. 4.5	4	$0 \leq j < 4$
	$s_{j,j',l_\Delta}^{(2)}$	Eq. 4.6	30	$0 \leq j \leq j' < 4$, $0 \leq l_\Delta < 3$
WPH	$S^{(1,1)}$	Eq. 4.7	3	$1 \leq j < 4$
	$S^{(0,0)}$	Eq. 4.8	3	$1 \leq j < 4$
	$S^{(0,1)}$	Eq. 4.9	3	$1 \leq j < 4$
$C(\ell)$		Eq. 4.1	20	linear bin, $\ell \in [20, 2020]$
$m_3(\theta)$		Eq. 4.10	12	smoothing scale $\theta \in [4\prime, 48\prime]$
$m_4(\theta)$		Eq. 4.11	12	smoothing scale $\theta \in [4\prime, 48\prime]$

Table 2. Summary of statistics used in the analysis, restricted to scales with $\ell \lesssim 2000$.

To build intuition about how each BCM parameter affects the statistics, we vary each parameter independently over its prior range defined in table 1, and plot the resulting statistics of $\kappa_{\mathbf{bary}}$ relative to those of $\kappa_{\mathbf{dmo}}$ in figure 3. Our final results in section 6 suggest that the three most important parameters to match and capture variations among different FLAMINGO feedback models are M_c, $\theta_{\rm ej}$, and η, which we highlight in figure 3. The effects of the remaining parameters are shown in figure 8 in appendix B. We find that the prior ranges induce a wide range of changes in the statistics — including both suppression and, in some cases, enhancement — relative to the fiducial variant's measurement.

5 Method

We want to demonstrate that the BCM can flexibly and accurately reproduce the wide range of baryonic feedback models in FLAMINGO, in terms of two-point and HOS and investigate what are the baryon parameter values that best fit the FLAMINGO simulations and how they vary as a function of hydrodynamical feedback strengths.

In principle, we could sample the best-fit baryon parameters in one step using a Markov Chain Monte Carlo (MCMC) — given the baryon parameters, we could generate baryonified κ maps using the DMO simulation, measure their statistics, and compare them to those from FLAMINGO. In practice, generating new baryonified maps and computing their statistics at each MCMC step is too time-consuming. Therefore, we proceed in three steps:

1. Build an emulator that predicts the statistics for the given baryon parameters
2. Use the emulator in an MCMC on each of the FLAMINGO simulations to find the best-fit baryon parameters
3. Once the MCMC chains have converged, we generate baryonified κ maps at the MCMC best fit, measure the suppression of the statistics, and use them for our final considerations.

Therefore, we use the emulator as a fast and efficient tool to refine our search in parameter space, rather than brute-forcing the problem by baryonifying the maps within the MCMC,

JCAP09(2025)073

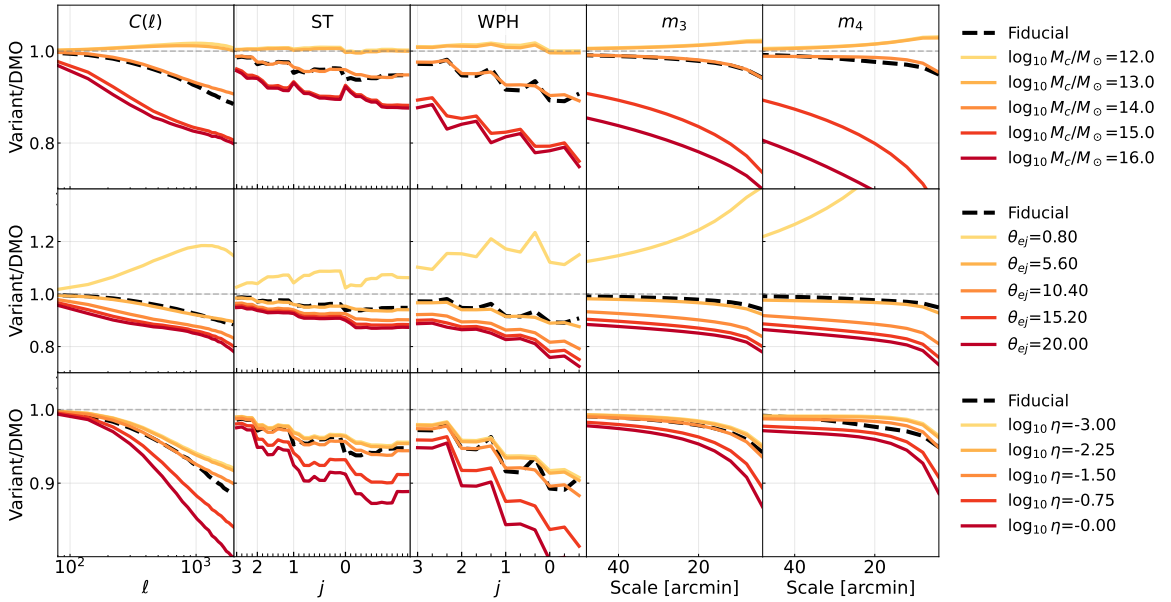

Figure 3. Summary statistics of the baryonified (colored) and the FLAMINGO fiducial variant's (black, dashed) convergence map, focusing on the first redshift bin. For the baryonified maps, a single BCM parameter is varied at a time. Rows correspond to different BCM parameters, and columns to different statistics. The y-axis shows the statistics of both the baryonified and FLAMINGO maps, relative to those of the DMO simulation with the same initial condition. On the x-axis, spatial scale decreases to the right. ST and WPH statistics are ordered by wavelet size (j), used here as a proxy for characteristic scale. The BCM flexibly generates a range of baryonic signatures in the statistics, clustering around the FLAMINGO results. The variation of other parameters in table 1 is shown in figure 8 in appendix B.

which would be computationally infeasible. This means that possible discrepancies between the suppression of the statistics in the FLAMINGO simulations and the one we obtain at the end of our three-step procedure can be attributed to two factors: (1) the accuracy of the emulator in reproducing the statistics compared to the full baryonification process, and (2) the intrinsic limitations of the baryonification method itself, even assuming a perfect emulator.

5.1 Emulator

We will refer to the collection of all summary statistics as $\mathcal{S}$. To accelerate the MCMC inference, we construct an emulator r, which predicts the ratio of the BCM-corrected summary statistics to its DMO counterpart given baryon parameters $\mathbf{b}$, i.e.,

$$r(\mathbf{b}) = \mathcal{S}_{\mathbf{bary}}(\mathbf{b})/\mathcal{S}_{\mathbf{dmo}},. \tag{5.1}$$

The numerator $\mathcal{S}_{\mathbf{bary}}(\mathbf{b})$ is computed by first baryonifying $\kappa_{\mathbf{dmo}}$ using the BCM (section 2) and then performing two-point and HOS measurements (section 4). Since $\kappa_{\mathbf{bary}}$ and $\kappa_{\mathbf{dmo}}$ share the same initial condition, the ratio eliminates shot noise and cosmic variance.

The training set includes $5,468$ $\kappa_{\mathbf{bary}}$ maps sampled in the ten-dimensional baryon parameter space. The prior ranges for the parameters are detailed in table 1 and are similar to those used in [56, 57]. In figures 3 and 8, we have shown how varying a single BCM parameter can produce suppression similar to those observed in the fiducial model.

JCAP09(2025)073

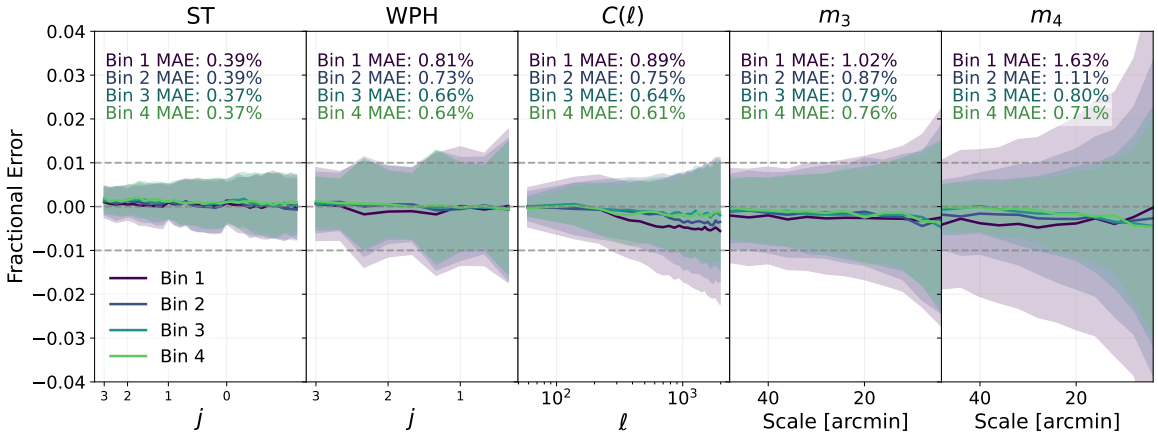

Figure 4. The emulator mean relative error, grouped by statistics and by redshift bins. The mean absolute (relative) error (MAE) is shown in text. The error is evaluated by averaging across the test set across the baryon parameter space. The emulator achieves sub-percent accuracy for all statistics besides the low redshift moments. The higher-order moment statistics are intrinsically nosier since it involves multiplications of random variables. We remind the reader that the emulator is used to efficiently explore the BCM parameter space when comparing to FLAMINGO statistics. However, the final assessment of how well the BCM reproduces a given FLAMINGO statistic is based on running the full baryonification pipeline at the emulator-derived best-fit parameters.

We also explicitly verify that, across the training samples, the baryonified maps exhibit baryonic signatures spanning the full range of variations seen in the different FLAMINGO feedback variants. For each statistic, we construct an ensemble of 10 neural networks via cross-validation. We use the mean and 1σ scatter of the ensemble predictions as the point estimate and uncertainty of $r(\mathbf{b})$. The emulator accuracy is shown in figure 4 as a function of scale and redshift, and generally decreases at smaller scales and lower redshifts, where baryonic feedback is more nonlinear and varied. The emulator achieves a mean absolute error below 1% for all statistics except the low-redshift moments, which are intrinsically noisier, as discussed in section 4.4. The emulator is implemented in `JAX` and supports automatic differentiation [105]. Details of the training set and emulator are provided in appendix C.

5.2 Sampling

For each FLAMINGO feedback variant with statistics $\mathcal{S}_{\mathbf{hydro}}$, we seek the best-fit baryon parameters by sampling the posterior $\mathcal{P}$

$$\mathcal{P}(\mathbf{b} \mid \mathcal{S}_{\mathbf{hydro}}) = \mathcal{G}\left(\frac{\mathcal{S}_{\mathbf{hydro}}}{\mathcal{S}_{\mathbf{dmo}}} - r(\mathbf{b}), \sigma^2 + \epsilon(\mathbf{b})^2\right) \tag{5.2}$$

where $\mathcal{G}$ is a Gaussian distribution parametrized by mean and variance, σ is a constant uncertainty per statistic and ϵ is the emulator uncertainty. We assume a flat prior on $\mathbf{b}$. As discussed above, the simulations are noiseless and the ratio statistics suppress the cosmic variance and shot noise. Therefore, in theory, the emulator uncertainty ϵ is the only source of error in $r(\mathbf{b})$. However, since ϵ is a function of $\mathbf{b}$, the parameter-dependent covariance makes the sampling very difficult in high-dimensional space. Therefore, we introduce a

JCAP09(2025)073

constant noise σ to promote the smoothness of the posterior surface to facilitate sampling. In particular, σ does not reflect real observational uncertainty. We find $\sigma = 0.04$ for ST, WPH, and $C(\ell)$ and $\sigma = 0.08$ for m_3 and m_4 make the chains converge robustly. Although we are putting less weight on the moments, we shall see that the final best-fit baryon parameters reproduce all statistics equally well. We leverage the gradient information provided by the emulator and use the `numpyro` implementation [106, 107] of the Hamiltonian Monte Carlo algorithm for efficient sampling [108, 109].

6 Results

6.1 The 10-parameter baryon correction model

We sample the posterior $\mathcal{P}(\mathbf{b} \mid \mathcal{S}_{\mathbf{hydro}})$ across the full 10-dimensional baryon parameter space for each of the FLAMINGO feedback variants. Since the marginal posterior distributions are approximately Gaussian, we use the medians — rather than the maximum a posteriori (MAP) values, which are less robust against sample variance — as our best-fit parameters $\hat{\mathbf{b}}$. The estimated $\hat{\mathbf{b}}$ values and their 1σ uncertainties are listed in table 4 in appendix D.

We then investigate how well the best-fit BCM reproduces the statistics in $\kappa_{\mathbf{flamingo}}$. Ideally, we would directly compare the emulator prediction $r(\hat{\mathbf{b}})$ with FLAMINGO. However, the emulator includes bias from both the neural network and the BCM itself. To isolate the latter, we directly construct $\kappa_{\mathbf{bary}}(\hat{\mathbf{b}})$, measure its statistics, and compare them to the FLAMINGO results. The emulator results agree with direct baryonification within expected emulator errors. For example, for FLAMINGO's fiducial variant, the relative error of the 10-parameter best-fit BCM is shown in figure 5 in dark red. The BCM accurately reproduces ST, WPH, $C(\ell)$, and moment statistics at all scales and redshifts to within 2%. The residual error is the largest at small scales and low redshifts, where baryonic feedback is strongest.

We analyze all 8 other FLAMINGO variants in the same way and show the residuals for the strongest (fgas-8σ) and weakest (fgas$+2\sigma$) FLAMINGO feedback variants in figure 9 and figure 10 in appendix E. The BCM can reproduce the statistics of different FLAMINGO variants equally well. For the strongest AGN feedback case (fgas-8σ), the BCM shows elevated errors of 3–4% in the first redshift bin of m_4 for $\theta < 20$ arcminutes, but otherwise maintains 2% accuracy. However, as discussed in section 4.4, higher-order moments are intrinsically noisier in both modeling and measurement. The observed m_4 discrepancy remains well below the statistical uncertainty of a DES-like survey [19] and is only detected at low significance in current DES Y3 data [37].

We have verified that a significant fraction of the remaining 2% residual at small scales can be alleviated by lowering the halo mass threshold to $10^{12.5}\,M_\odot/h$. This test was performed using the best-fit of the baryonification emulator on the Fiducial simulation, and should be considered an approximate check. A formal evaluation would require regenerating the maps, retraining the emulator, and refitting the model using $10^{12.5}\,M_\odot/h$ as the threshold — a process we do not pursue due to its substantial computational cost. Nonetheless, we note that adopting a threshold of $10^{13}\,M_\odot/h$ still yields sub-2% precision.

Thus far, we have investigated the BCM at the fiducial cosmology. While both $\mathcal{S}_{\mathbf{hydro}}$ and $\mathcal{S}_{\mathbf{dmo}}$ depend on cosmology and astrophysics, these dependencies are approximately

JCAP09(2025)073

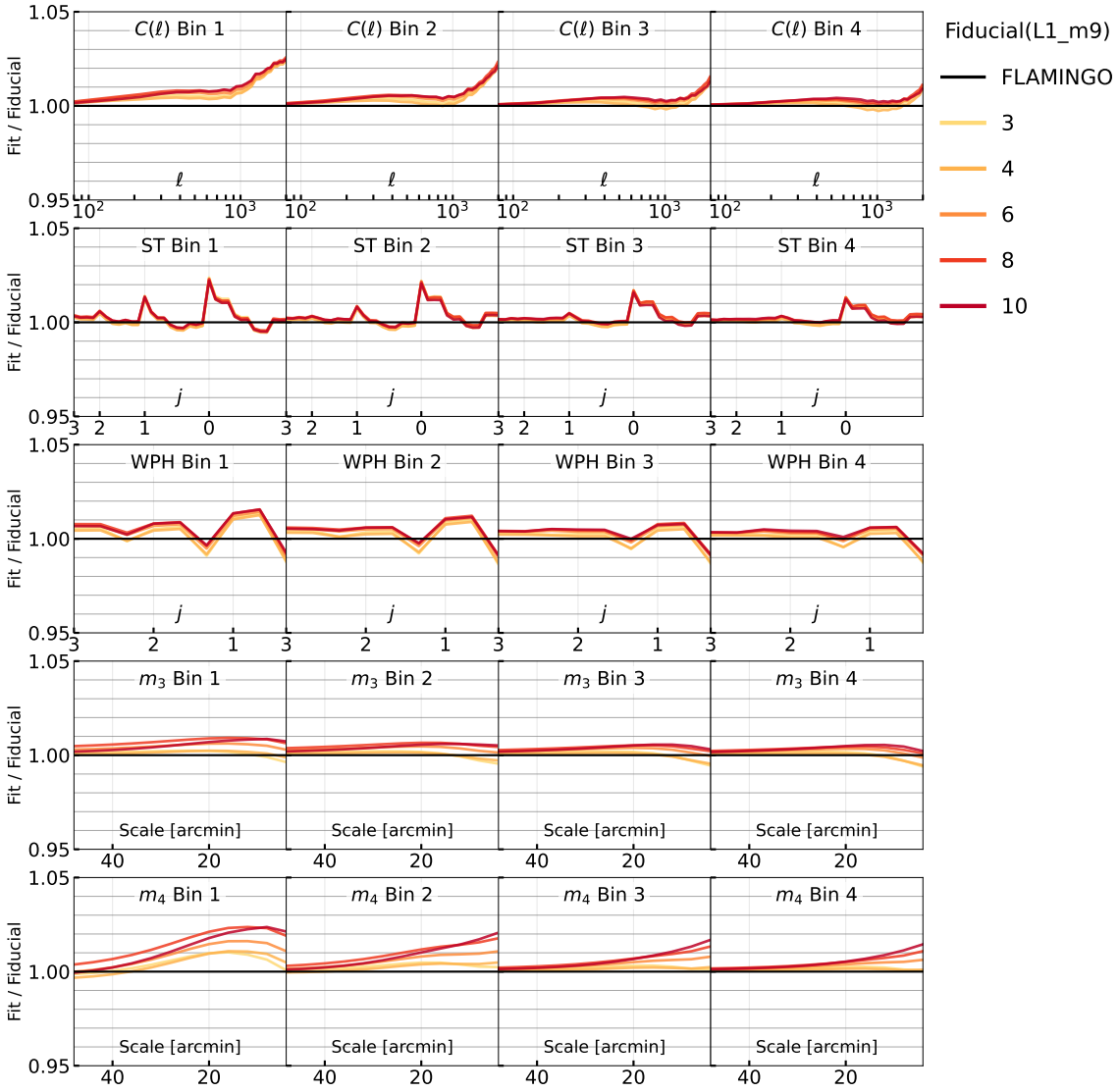

Figure 5. The residuals error of the BCM when fitted to the FLAMINGO fiducial variant. Rows correspond to different statistics, and columns to redshift bins. The y-axis shows the ratio between the statistics of the best-fit baryonified map to that of the fiducial variant. On the x-axis, spatial scales decreases to the right. ST and WPH are ordered by the wavelet size (j) as a proxy for their characteristic scale. Colors indicate the number of free parameters in the BCM fit. The shaded region marks $\pm 2\%$ for visual reference. For all n-parameter models, the BCM flexibly reproduces the baryonic features in the FLAMINGO fiducial variant simulation to percent level. The prediction is robust even with only 3 degrees of freedoms. Differences in accuracy between the models stem from how we estimate the best-fit parameter, as discussed in section 6.2. The residuals for the strongest (fgas-8σ) and weakest (fgas$+2\sigma$) feedback variants are shown in figure 9 and figure 10 in appendix E.

JCAP09(2025)073

Variant	$\log_{10} M_c/M_\odot$	$\log_{10}\eta$	$\theta_{\rm ej}$
Fiducial(L1_m9)	13.9 ± 0.19	-1.51 ± 0.39	4.80 ± 0.65
Jet	13.7 ± 0.16	-1.87 ± 0.39	$6.23^{+0.89}_{-0.75}$
Jet_fgas-4σ	14.0 ± 0.16	$-1.02^{+0.21}_{-0.25}$	5.94 ± 0.61
fgas$+2\sigma$	13.7 ± 0.16	-1.95 ± 0.38	$5.05^{+0.68}_{-0.59}$
fgas-2σ	14.0 ± 0.22	$-1.20^{+0.34}_{-0.35}$	4.85 ± 0.59
fgas-4σ	14.0 ± 0.19	$-0.983^{+0.23}_{-0.29}$	5.05 ± 0.54
fgas-8σ	14.0 ± 0.14	$-0.885^{+0.17}_{-0.18}$	5.40 ± 0.52
M*$-\sigma$	14.0 ± 0.21	$-1.17^{+0.33}_{-0.37}$	4.69 ± 0.55
M*$-\sigma$_fgas-4σ	14.0 ± 0.15	-0.852 ± 0.2	5.05 ± 0.5

Table 3. The best-fit BCM parameters obtained from the joint fitting of all HOS to different FLAMINGO variants. These parameter combinations yield HOS that agree with FLAMINGO measurements to 2%. Other BCM parameters are fixed to $\nu_{M_c} = 0$, $\nu_{\theta_{\rm ej}} = 0$, $\gamma = 2.60$, $\log_{10}\theta_{co} = -1.00$, $\log_{10}\eta_\delta = -0.28$, $\delta = 14.0$, and $\mu_\beta = 1.60$. The uncertainty is the 16% and 84% credible interval of our fitting method. See table 4 in appendix D for the best-fit values for all n-parameter models.

factorizable [26, 110]. As a result, the cosmology dependence largely cancels in the ratio $\mathcal{S}_{\rm hydro}/\mathcal{S}_{\rm dmo}$, which the BCM models. Using the low-S_8 variants (LS8, LS8_fgas-8σ) of the FLAMINGO suite, we show in appendix F that this ratio varies by less than 1% compared to the fiducial cosmology even for the strong feedback model (fgas-8σ). Since the BCM (and thus the emulator) predicts the suppression ratio solely from the baryonic parameters, the recovered LS8 BCM parameters are consistent with those from the fiducial cosmology. The baryonic suppression has a stronger dependence on the baryon fraction, Ω_b/Ω_m [e.g. 56] but we leave a detailed analysis of this dependence to future work due to the computational cost of training separate emulators for each cosmology.

To summarize, for $\ell \leq 2000$, the 10-parameter BCM is flexible enough to reproduce the range of baryonic signatures in two-point and HOS.

6.2 Baryon correction model with restricted degrees of freedoms

After validating that the 10-parameter BCM accurately reproduces the statistics, we ask whether similar success can be achieved with significantly fewer degrees of freedom. This simplification has three advantages. First, in real survey analyses, using fewer baryon parameters reduce the impact on the constraining power of key cosmological parameters. Second, in simulation-based inference, it lowers the number of simulations required to train the neural posterior estimator, thus reducing computational cost. Third, the 10-dimensional $\mathbf{b}$ space contains significant degeneracies; fewer parameters improves the interpretability how $\mathbf{b}$ vary across FLAMINGO feedback variants.

We consider 4 restricted versions of the BCM, with 8, 6, 4, and 3 parameters. In general, at each step, we fix parameters whose best-fit values are consistent across all feedback variants. The 8-parameter model fixes the redshift scalings ν_{M_c} and $\nu_{\theta_{\rm ej}}$ to zero, consistent with their posteriors in the 10-parameter fits. The 6-parameter model further fixes $\gamma = 2.6$

JCAP09(2025)073

and $\log_{10}\theta_{co} = -1$. For the 4-parameter model, we also fix $\log_\delta \eta_\delta = 0.28$ and $\delta = 14$. Finally, for the 3-parameter model, we additionally fix $\mu_\beta = 1.6$. The remaining free parameters in the 3-parameter model are $\log_{10} M_c$, $\log_{10}\eta$, and $\theta_{\rm ej}$. For each n-parameter model and feedback variant, we tabulate $\hat{\mathbf{b}}$ in table 4 in appendix D, with fixed parameters shown in red. We also summarize $\hat{\mathbf{b}}$ values for the 3-parameter model in table 3. As with the 10-parameter case, we validate the best-fit BCMs by directly constructing $\kappa_{\mathbf{bary}}$ and measuring their statistics. Residuals for the fiducial, strong feedback (fgas-8σ), and weak feedback (fgas$+2\sigma$) variants are shown in figures 5, 9, and 10, respectively.

Although we have dramatically reduced the number of parameters, all n-parameter models faithfully reproduce all statistics across scales, redshifts, and feedback variants. Notably, the 3-parameter model often fits the statistics better than the 10-parameter model. This is because our best estimate, $\hat{\mathbf{b}}$, is defined as the median of the marginal posterior. In high-dimensional spaces, $\hat{\mathbf{b}}$ can deviate from the MAP due to projection over non-Gaussian volumes. As the number of parameters decreases, the posterior becomes more Gaussian-like, and $\hat{\mathbf{b}}$ approaches the MAP, leading to improved fits. In other words, were we to have a less noisy emulator, we could potentially use the MAP as $\hat{\mathbf{b}}$ and obtain even higher-quality fits for the 10-parameter model.

6.3 Baryon parameters

Having demonstrated that the BCM can flexibly reproduce the baryonic signatures across a range of FLAMINGO feedback variants, we now examine the best-fit baryon parameters themselves. When comparing with other results in the literature, we emphasize that our modeling choices — such as the specific implementation of the BCM and the set of free parameters — are in some aspects different from those adopted in other studies, and that our results are entirely based on simulations. In this context, consistency at the parameter level with earlier work is not required for our approach to be effective, as success is defined by how well it reproduces the suppression of the lensing statistics in the FLAMINGO simulations, rather than by agreement in individual parameter values. Nonetheless, the inferred $\hat{\mathbf{b}}$ is broadly consistent with earlier results based on both simulations and observations [56, 59, 62]. This qualitative agreement reinforces that our model is not overfitting to a narrow region of parameter space and is instead capturing physically meaningful baryonic effects.

Focusing on the 10-parameter model, all variants are consistent with $\nu_{M_c} = 0$ and $\nu_{\theta_{\rm ej}} = 0$, suggesting minimal sensitivity of our measurements to the mass dependence of the gas ejection radius and characteristic slope. This lack of sensitivity is consistent with the forecasted constraints from weak lensing presented in [57]. Interestingly, all feedback variants prefer a high δ, which corresponds to a sharp drop off in the gas density profile past $r > \theta_{\rm ej} R_{\rm 200c}$ [56, 57, 62]. To investigate if different values of δ affect the best-fit statistics, we rerun the 4- and 3-parameter models, setting $\delta = 7$ and $\mu_\beta = 2$ while keeping the other fixed parameters the same as in section 6.2. The BCM still achieves a 2% fit across all statistics and feedback variants, although the best-fit $\theta_{\rm ej}$ lowers from 5 to 3. This is expected as δ and $\theta_{\rm ej}$ both control the radial extent of the gas profile and are therefore quite negatively correlated. Next, we investigate how much the estimated $\hat{\mathbf{b}}$ varies across the n-parameter models. Figure 6 shows $\hat{\mathbf{b}}$ as a function of feedback variant for the 6-, 4-, and 3-parameter

JCAP09(2025)073

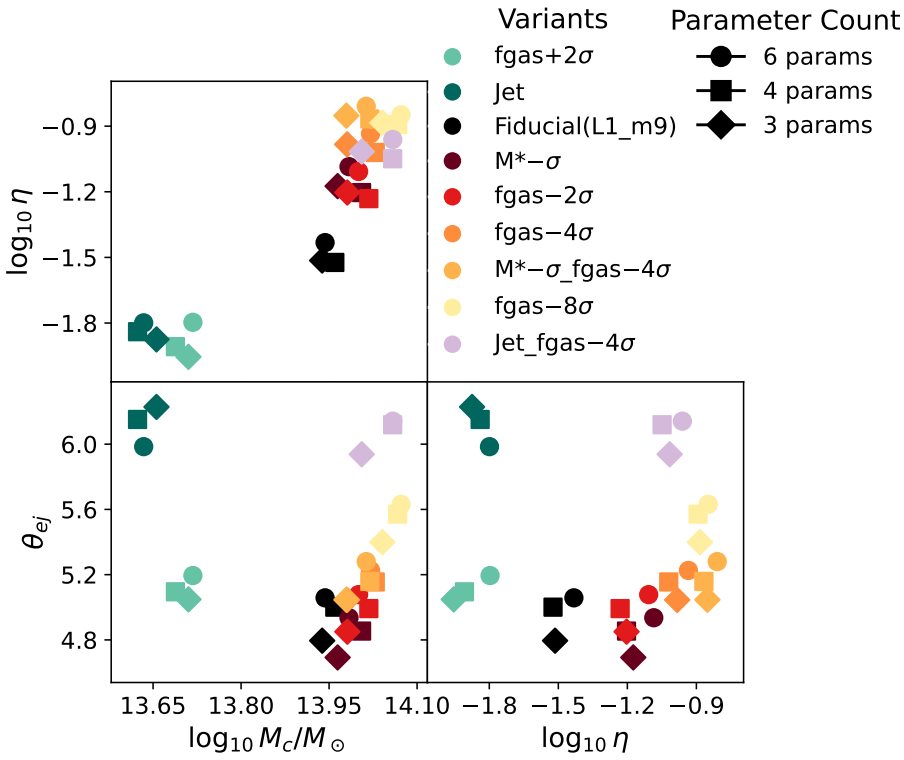

Figure 6. The inferred baryon parameter $\hat{\mathbf{b}}$ from different FLAMINGO feedback variants. Markers represent inference using different degrees of freedom. The color map labels the feedback variants and inherits from figure 2. In the legend, the feedback strength approximately increases from top to bottom, with the strongest (weakest) feedback model suppression being the fgas-8σ (fgas$+2\sigma$) model. For each feedback variant, the clustering of the points demonstrate the BCM is robust against removing many nuisance parameter dependencies.

models. The constraints on the remaining free parameters do not vary significantly as we reduce the number of parameters. This demonstrates that the BCM's weak lensing predictions are robust against removing many nuisance parameters and the BCM allows us to track how the phenomenological baryon parameters change as a function of FLAMINGO feedback models. The best-fit parameters of the different models show some degree of clustering in the 3D parameter space, indicating that, if sufficient statistical power is achieved by observational probes, it may be possible to distinguish between feedback models using real data. Concerning the physical interpretation of the position of the models in the parameter space, we know that $\log_{10} M_c$, $\log_{10} \eta$, and $\theta_{\rm ej}$ all scale positively with feedback strength. Indeed, we find that the strongest feedback model (fgas-8σ) lies in the upper range of each of the three parameters, while the weakest feedback model (fgas$+2\sigma$ weak) occupies the lowest values in each dimension. Beyond a qualitative ranking of suppression strength, the interplay between the three parameters and the resulting impact on the lensing statistics is non-trivial. Each parameter modulates the suppression differently and in a scale-dependent way, leading to best-fit values that are scattered across parameter space in a complex pattern.

JCAP09(2025)073

7 Conclusion

Data from current and upcoming weak lensing surveys are probing cosmological structures at increasingly small spatial scales. To extract both cosmological and astrophysical information in this regime robustly and optimally, we require models that can accurately capture the full range of baryonic feedback effects allowed by theory and observed in state-of-the-art hydrodynamical simulations.

In this work, we investigate the map-level baryonification approach from [57], which efficiently and flexibly models baryonic effects in weak lensing convergence maps. By calibrating the baryon parameters against FLAMINGO hydrodynamical simulations, we demonstrate that this baryon correction model (BCM) accurately reproduces baryonic impacts on various convergence field statistics, including the two-point power spectrum, scattering transform coefficients (ST), wavelet phase harmonics (WPH), and third- and fourth-order moments. This agreement holds across scales $\ell < 2000$ and throughout the redshift range of DES-like surveys. We construct an emulator to predict BCM-generated statistics efficiently, then sample baryon parameters by comparing emulator predictions with measurements from different FLAMINGO feedback variants. Testing BCM versions with 10, 8, 6, 4, and 3 free parameters, we find all models can reproduce statistics within $< 2\%$ accuracy with accuracy improving at larger scales and higher redshifts. A minor exception is the fourth moment, where, for models with more than 3 free parameters, the residual error approaches 3% around $10\prime$, partially due to the increased uncertainty in our fitting procedure.

These results have several implications for future weak lensing analyses. First, this shows that the BCM is robust for a wide range of weak lensing statistics and can be embedded in simulation-based inference frameworks to marginalize over baryonic uncertainties or directly constrain baryon parameters. Reduced-parameter versions of the BCM further improve efficiency by lowering the number of simulation needed and sampling costs. Second, our work can be extended to test the robustness of other tracers and probes, such as cross-correlations with large-scale structure observables like galaxy clustering or thermal Sunyaev-Zel'dovich maps. Joint analyses with these additional datasets may offer tighter constraints on baryonic physics.

Although our results are based on the DES Y3 redshift distributions, they are broadly applicable to other current weak lensing surveys due to similar source redshift distributions [111–113]. They are also relevant for upcoming surveys like LSST, Euclid, and Roman, whose higher-redshift distributions may allow our models to extend to even smaller angular scales due to weaker baryonic feedback at early times. The ability to model baryonic physics in this regime opens new opportunities for extracting both cosmological and astrophysical information from small-scale data. Future work could extend this framework to alternative summary statistics, such as voids, peaks, or CNN-based approaches [e.g., 16, 114]. Our results highlight the promise of map-level baryonification — without relying on 3D particle snapshots — as a computationally efficient and physically grounded approach for incorporating baryonic feedback in weak lensing analyses. As upcoming surveys push to smaller scales, such methods will be essential for maximizing constraining power while remaining robust to astrophysical systematics.

JCAP09(2025)073

Acknowledgments

AZ is supported by Jane Street through the Jane Street Graduate Research Fellowship. MG is supported by funds provided by the Kavli Institute for Cosmological Physics at the University of Chicago through an endowment from the Kavli Foundation. DA is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1746045. This work was supported by FermiForward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (http://www.dirac.ac.uk/). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

Data availability. The BaryonForge pipeline used in this work is publicly available at https://github.com/DhayaaAnbajagane/BaryonForge. The FLAMINGO simulations are available upon request by contacting Joop Schaye at schaye@strw.leidenuniv.nl. The total matter power spectra from the FLAMINGO simulations have been publicly released and are available at https://flamingo.strw.leidenuniv.nl.

A Scale cut

We empirically determine the characteristic scale of each ST and WPH statistic and enforce a consistent scale cut with respect to the power spectrum. For each statistic $\mathcal{S}$, we smooth the $N_{\rm side} = 1024$ $\kappa_{\rm dmo}$ maps with low-pass filters at different $\ell_{\rm max}$ values and measure the statistics ($\mathcal{S}_\ell$), normalized by the unsmoothed map's statistics ($\mathcal{S}_\infty$). The results for ST and WPH are shown in the left and right panels of figure 7, respectively. For a given statistic, if $\mathcal{S}_\ell/\mathcal{S}_\infty \ll 1$, then it contains information beyond ℓ, and vice versa. We exclude statistics where $\mathcal{S}_\ell/\mathcal{S}_\infty < 0.5$ at $\ell_{\rm max} = 2000$. All 34 ST coefficients per redshift bin remain within our analysis scales. For WPH, the three $j = 0$ coefficients (shown in red) fail the scale cut criterion and are removed from the analysis.

B Statistics as a function of baryon parameters

Here we show how the summary statistics vary with the remaining BCM parameters listed in table 1, excluding M_c, $\theta_{\rm ej}$, and η, which are highlighted in the main text (figure 3). Each parameter is varied individually across its prior range, and the resulting baryonified statistics are compared to those from the corresponding DMO simulation. These results show the flexibility of the BCM in reproducing a wide range of baryonic effects around the hydrodynamical predictions.

C Emulator

We build a neural network-based emulator to map the baryon parameters to the tomographic ST, WPH, $C(\ell)$ and the m_3 and m_4 statistics. Each type of statistic is modeled independently.

JCAP09(2025)073

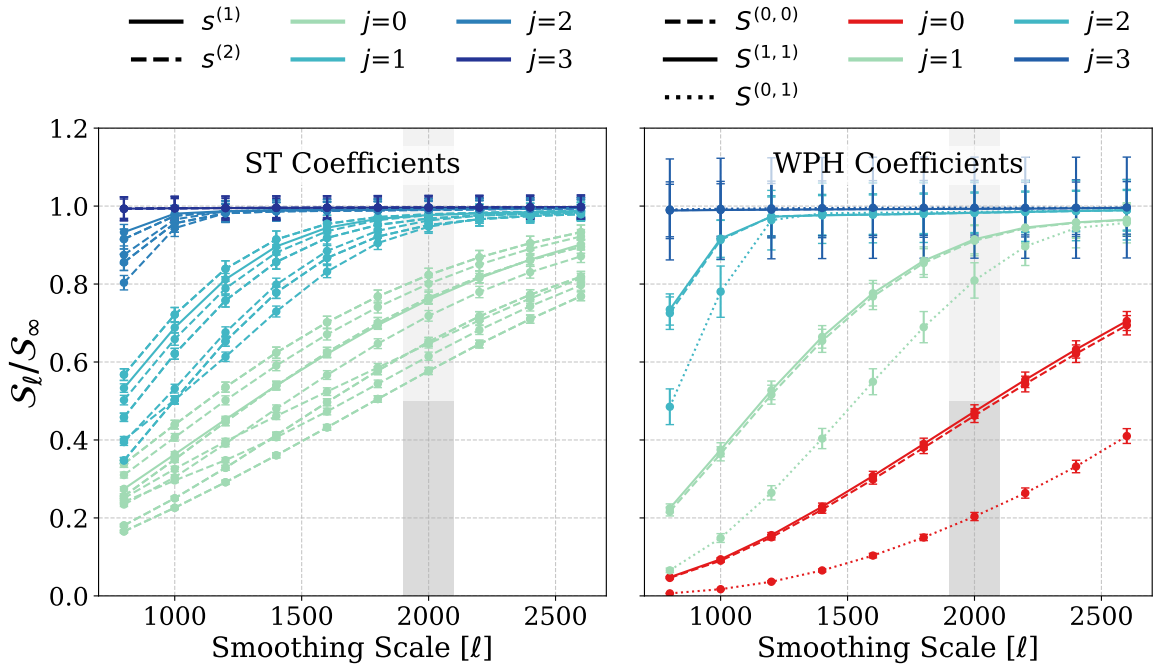

Figure 7. The effect of smoothing scale on the two kinds of wavelet coefficients, ST and WPH. Given a statistics $\mathcal{S}$, we define $\mathcal{S}_\ell$ as the statistics measured on κ maps that are low-pass filtered (with a step function) at different ℓ. The statistics measured on the smoothed map is called $\mathcal{S}_\infty$. We test this idea on $\kappa_{\mathbf{dmo}}$, and plot the relative suppression $\mathcal{S}_\ell/\mathcal{S}_\infty$ as a function of smoothing scale ℓ. The line and the error bar represent the mean and the standard error of $\mathcal{S}$ measured on different patches across the full sky. We further color code the lines using the characteristic scale of the wavelet (higher j equals larger wavelets). If $\mathcal{S}_\ell/\mathcal{S}_\infty \ll 1$, then $\mathcal{S}$ contains information beyond ℓ. The $j = 0$ WPH coefficients (shown in red) are suppressed more than 50% at $\ell = 2000$ (shaded region), and hence removed from subsequent analyses.

We first standardize the input parameters and then use three dense hidden layers following the architectures: (512, 256, 256) for ST, (512, 256, 128) for WPH, (512, 256, 128) for $C(\ell)$ statistics, and (512, 256, 128) for the concatenated m_3 and m_4 data vector. Each hidden layer is followed by a leaky ReLU activation function ($\alpha = 0.1$) [115] and a 3% dropout layer to prevent over-fitting [116]. For robustness, we train 10 independent cross-validated models for each type of statistics, resulting in a total ensemble of 40 neural networks. We use 95% of the data for the 10-fold cross-validation and use the rest as the test set for the performance evaluation as shown in figure 4. During training, we use the mean absolute error (MAE) as the loss function and the Adam optimizer with weight decay for l_2 regularization [117]. The emulator implements a dual early stopping strategy to prevent over-fitting: (1) a global early stopping mechanism with a patience of 80 iterations that monitors validation loss to terminate training when no improvement is observed, and (2) a learning rate scheduler that starts with a learning rate of 10^{-3} and reduces it by a factor of 0.5 after 5 epochs without improvement, with a minimum threshold of 10^{-6}. Since each statistic is predicted 10 times by the cross-validated models, our emulator not only provides a point estimate (given by the ensemble mean) but also uncertainty estimates (given by the ensemble variance). The entire

JCAP09(2025)073

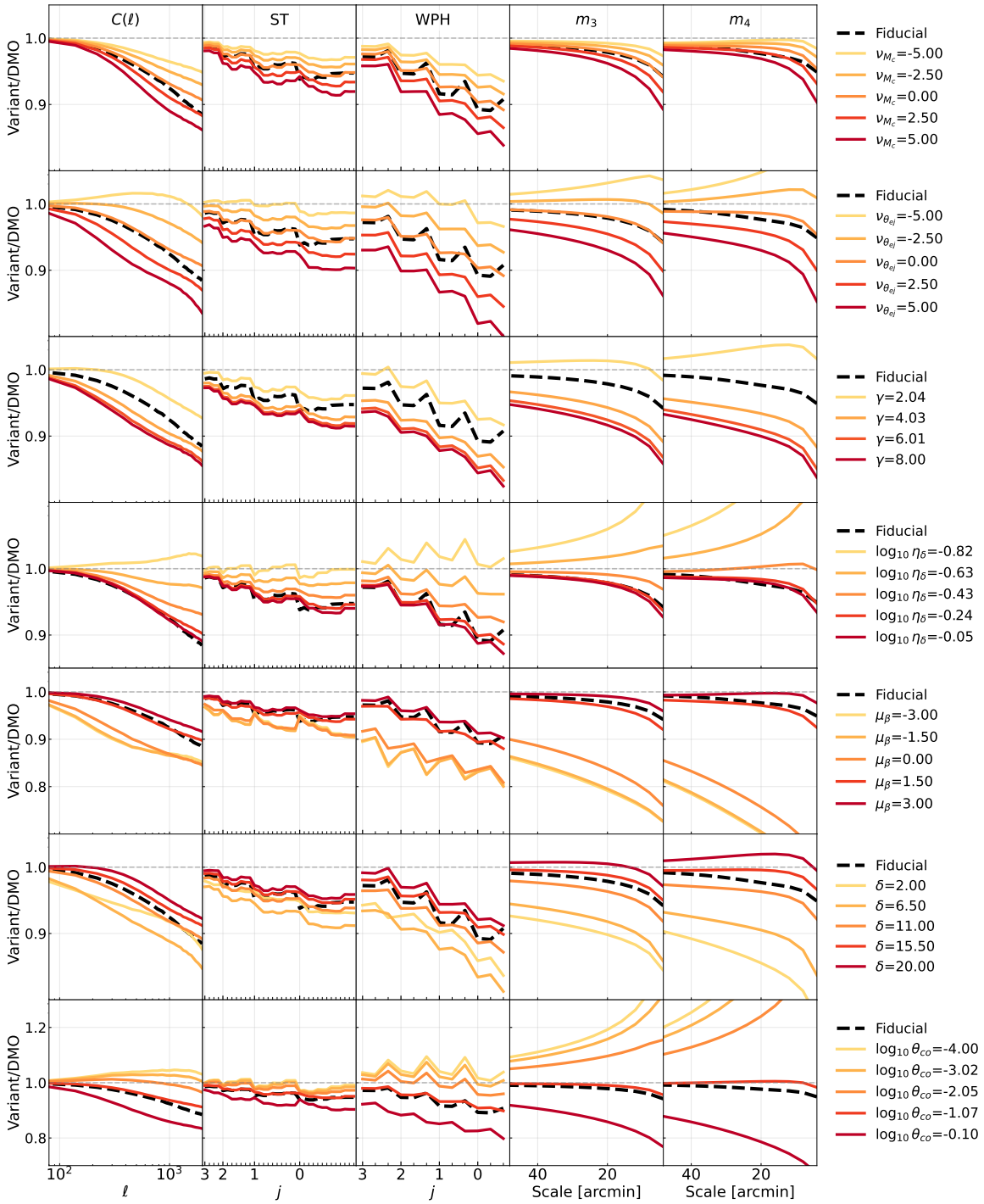

Figure 8. Same as figure 3, but for the remaining BCM parameters not shown in the main text. Colored curves show the summary statistics of baryonified convergence maps with one parameter varied at a time, while black dashed curves show the corresponding statistics from the FLAMINGO simulations.

JCAP09(2025)073

implementation supports automatic differentiation through `JAX`, enabling gradient-based optimization and sampling, as well as sensitivity analysis of the statistics with respect to the baryon parameters [105].

We used the baryonification model to generate 6695 $\kappa_{\mathbf{bary}}(\mathbf{b})$ maps and measured their statistics $\mathcal{S}_{\mathbf{bary}}$, where the 10-dimensional $\mathbf{b}$ is sampled using a Sobol sequence. Since some $\mathcal{S}_{\mathbf{bary}}$ show an amplification rather than a suppression as observed in all FLAMINGO measurements, we remove $\mathcal{S}_{\mathbf{bary}}$ where $C(\ell)$ is on average amplified by 10% across all linear bins. The final training set includes 5468 $\mathcal{S}_{\mathbf{bary}}$. Since the training set is small, the emulator has high outlier rate near parameter boundaries. Therefore, we truncate parameter domain boundaries where the absolute error averaged across all statistics is above 1%. The most significant truncation occurs for $\gamma < 0.9$. We later verify these restrictions do not impact the posteriors of the MCMC chains.

D Best-fit baryon parameters

JCAP09(2025)073

Variant	ν_{M_c}	$\nu_{\theta_{\rm ej}}$	γ	$\log_{10}\theta_{co}$	$\log_{10}\eta_\delta$	δ	μ_β	$\log_{10}M_c/M_\odot$	$\log_{10}\eta$	$\theta_{\rm ej}$
Fiducial(L1_m9)(10)	$1.48^{+2.0}_{-2.4}$	$-0.727^{+1.1}_{-0.99}$	$2.61^{+1.7}_{-0.81}$	-1.06 ± 0.19	$-0.247^{+0.12}_{-0.16}$	13.7 ± 3.5	1.81 ± 0.53	13.9 ± 0.25	-1.57 ± 0.43	$5.36^{+2.4}_{-1.7}$
Fiducial(L1_m9)(8)	0	0	$2.64^{+1.8}_{-0.85}$	-1.04 ± 0.18	$-0.255^{+0.13}_{-0.18}$	$14.4^{+3.1}_{-3.7}$	1.77 ± 0.56	13.9 ± 0.18	$-1.49^{+0.48}_{-0.44}$	$4.85^{+1.8}_{-1.3}$
Fiducial(L1_m9)(6)	0	0	2.60	-1.00	$-0.282^{+0.13}_{-0.17}$	$15.1^{+2.8}_{-3.9}$	$1.87^{+0.56}_{-0.48}$	13.9 ± 0.18	-1.43 ± 0.46	$5.06^{+0.91}_{-1.1}$
Fiducial(L1_m9)(4)	0	0	2.60	-1.00	-0.28	14.0	$1.92^{+0.5}_{-0.42}$	14.0 ± 0.16	$-1.52^{+0.37}_{-0.41}$	5.00 ± 0.7
Fiducial(L1_m9)(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	13.9 ± 0.19	-1.51 ± 0.39	4.80 ± 0.65
Jet(10)	$0.697^{+2.4}_{-2.7}$	$-0.227^{+1.1}_{-1.1}$	$2.86^{+2.0}_{-0.98}$	$-0.959^{+0.22}_{-0.21}$	-0.305 ± 0.15	$13.5^{+3.8}_{-3.9}$	$1.50^{+0.72}_{-0.55}$	13.6 ± 0.32	-1.89 ± 0.41	$5.25^{+2.6}_{-1.7}$
Jet(8)	0	0	$2.92^{+2.0}_{-1.0}$	-0.956 ± 0.2	-0.313 ± 0.15	$14.0^{+3.6}_{-4.2}$	$1.45^{+0.72}_{-0.58}$	13.7 ± 0.28	-1.86 ± 0.44	$5.01^{+2.1}_{-1.5}$
Jet(6)	0	0	2.60	-1.00	$-0.282^{+0.13}_{-0.13}$	$14.7^{+3.1}_{-4.1}$	$1.45^{+0.76}_{-0.54}$	13.6 ± 0.27	$-1.80^{+0.38}_{-0.41}$	5.98 ± 1.4
Jet(4)	0	0	2.60	-1.00	-0.28	14.0	$1.49^{+0.74}_{-0.5}$	13.6 ± 0.26	$-1.84^{+0.38}_{-0.42}$	6.15 ± 0.98
Jet(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	13.7 ± 0.16	-1.87 ± 0.39	$6.23^{+0.89}_{-0.75}$
Jet_fgas-4σ(10)	$1.79^{+1.9}_{-2.3}$	$-0.785^{+1.2}_{-1.0}$	$2.70^{+1.6}_{-0.77}$	-1.08 ± 0.15	$-0.249^{+0.13}_{-0.23}$	$13.5^{+3.4}_{-3.2}$	1.93 ± 0.39	14.0 ± 0.2	$-1.13^{+0.34}_{-0.37}$	$6.84^{+2.8}_{-2.0}$
Jet_fgas-4σ(8)	0	0	$2.86^{+1.6}_{-0.84}$	-1.07 ± 0.13	$-0.295^{+0.16}_{-0.34}$	$14.3^{+3.2}_{-3.5}$	1.92 ± 0.37	14.1 ± 0.16	$-0.99^{+0.32}_{-0.35}$	$5.78^{+1.9}_{-1.3}$
Jet_fgas-4σ(6)	0	0	2.60	-1.00	$-0.328^{+0.17}_{-0.27}$	$14.9^{+2.8}_{-3.5}$	$1.94^{+0.36}_{-0.32}$	14.1 ± 0.14	$-0.961^{+0.24}_{-0.31}$	6.14 ± 0.98
Jet_fgas-4σ(4)	0	0	2.60	-1.00	-0.28	14.0	$1.96^{+0.35}_{-0.31}$	14.1 ± 0.13	$-1.05^{+0.21}_{-0.26}$	$6.12^{+0.69}_{-0.62}$
Jet_fgas-4σ(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.16	$-1.02^{+0.21}_{-0.25}$	5.94 ± 0.61
fgas$+2\sigma$(10)	$0.761^{+2.4}_{-2.6}$	$-0.328^{+1.1}_{-1.1}$	$2.77^{+2.0}_{-0.9}$	-1.07 ± 0.22	-0.291 ± 0.15	$13.7^{+3.6}_{-4.0}$	$1.58^{+0.71}_{-0.59}$	13.7 ± 0.29	-1.88 ± 0.43	$5.08^{+2.4}_{-1.7}$
fgas$+2\sigma$(8)	0	0	$2.75^{+1.9}_{-0.93}$	$-1.05^{+0.2}_{-0.19}$	-0.302 ± 0.15	$14.1^{+3.4}_{-3.9}$	$1.57^{+0.72}_{-0.59}$	13.7 ± 0.24	-1.85 ± 0.43	$4.82^{+2.1}_{-1.4}$
fgas$+2\sigma$(6)	0	0	2.60	-1.00	$-0.317^{+0.13}_{-0.13}$	$15.0^{+3.2}_{-3.9}$	$1.66^{+0.75}_{-0.64}$	13.7 ± 0.25	-1.80 ± 0.43	$5.19^{+1.1}_{-1.1}$
fgas$+2\sigma$(4)	0	0	2.60	-1.00	-0.28	14.0	$1.66^{+0.79}_{-0.61}$	13.7 ± 0.25	-1.91 ± 0.4	5.10 ± 0.73
fgas$+2\sigma$(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	13.7 ± 0.16	-1.95 ± 0.38	$5.05^{+0.68}_{-0.59}$
fgas-2σ(10)	$1.69^{+2.0}_{-2.2}$	$-0.802^{+1.0}_{-0.95}$	$2.59^{+1.5}_{-0.73}$	-1.06 ± 0.16	$-0.237^{+0.12}_{-0.19}$	$13.8^{+3.5}_{-3.3}$	1.84 ± 0.45	13.9 ± 0.22	$-1.31^{+0.43}_{-0.39}$	$5.53^{+2.3}_{-1.6}$
fgas-2σ(8)	0	0	$2.73^{+1.4}_{-0.88}$	-1.03 ± 0.15	$-0.267^{+0.14}_{-0.29}$	$14.5^{+3.2}_{-3.5}$	$1.83^{+0.45}_{-0.43}$	14.0 ± 0.18	$-1.17^{+0.41}_{-0.43}$	$4.86^{+1.8}_{-1.2}$
fgas-2σ(6)	0	0	2.60	-1.00	$-0.295^{+0.15}_{-0.23}$	$15.2^{+2.8}_{-3.7}$	$1.86^{+0.45}_{-0.38}$	14.0 ± 0.17	$-1.11^{+0.33}_{-0.42}$	$5.08^{+0.83}_{-0.93}$
fgas-2σ(4)	0	0	2.60	-1.00	-0.28	14.0	$1.88^{+0.4}_{-0.36}$	14.0 ± 0.17	$-1.23^{+0.32}_{-0.34}$	4.99 ± 0.63
fgas-2σ(3)	0	0	2.60	-1.00	-0.28	14.0	1.60	14.0 ± 0.22	$-1.20^{+0.34}_{-0.35}$	4.85 ± 0.59

Continued on next page

Table 4. The best-fit BCM parameters obtained from the joint fitting of all HOS to different FLAMINGO variants. These parameter combinations yield HOS that agree with FLAMINGO measurements to 2%. The integers following the feedback variant names are the number of free parameters. The fixed parameters are highlighted in red. The uncertainty is the 16% and 84% credible interval.

Variant	ν_{M_c}	$\nu_{\theta_{\rm ej}}$	γ	$\log_{10}\theta_{co}$	$\log_{10}\eta_\delta$	δ	μ_β	$\log_{10} M_{\rm c}/M_\odot$	$\log_{10}\eta$	$\theta_{\rm ej}$
fgas−4σ(10)	$1.68^{+2.0}_{-2.1}$	$-0.822^{+1.0}_{-0.95}$	$2.64^{+1.4}_{-0.73}$	-1.04 ± 0.15	$-0.24^{+0.13}_{-0.24}$	$13.8^{+3.3}_{-3.1}$	1.89 ± 0.41	14.0 ± 0.2	$-1.12^{+0.35}_{-0.37}$	$5.64^{+2.1}_{-1.5}$
fgas−4σ(8)	0	0	$2.81^{+1.4}_{-0.83}$	-1.02 ± 0.13	$-0.293^{+0.17}_{-0.34}$	$14.4^{+3.1}_{-3.4}$	$1.86^{+0.4}_{-0.37}$	14.0 ± 0.16	$-0.964^{+0.31}_{-0.36}$	$4.86^{+1.5}_{-1.0}$
fgas−4σ(6)	0	0	2.60	−1.00	$-0.309^{+0.16}_{-0.25}$	$15.1^{+2.8}_{-3.6}$	$1.85^{+0.39}_{-0.34}$	14.0 ± 0.15	$-0.935^{+0.25}_{-0.32}$	5.23 ± 0.85
fgas−4σ(4)	0	0	2.60	−1.00	−0.28	14.0	$1.88^{+0.37}_{-0.31}$	14.0 ± 0.15	$-1.02^{+0.22}_{-0.28}$	5.16 ± 0.58
fgas−4σ(3)	0	0	2.60	−1.00	−0.28	14.0	1.60	14.0 ± 0.19	$-0.983^{+0.23}_{-0.29}$	5.05 ± 0.54
fgas−8σ(10)	$1.51^{+2.0}_{-2.1}$	$-0.753^{+1.1}_{-0.93}$	$2.74^{+1.4}_{-0.76}$	-1.04 ± 0.13	$-0.275^{+0.16}_{-0.32}$	14.0 ± 3.2	1.92 ± 0.38	14.0 ± 0.19	$-0.93^{+0.28}_{-0.3}$	$5.94^{+2.4}_{-1.6}$
fgas−8σ(8)	0	0	$2.96^{+1.4}_{-0.85}$	-1.03 ± 0.12	$-0.323^{+0.18}_{-0.36}$	$14.6^{+3.0}_{-3.4}$	1.93 ± 0.36	14.1 ± 0.14	$-0.857^{+0.26}_{-0.24}$	$5.12^{+1.5}_{-1.0}$
fgas−8σ(6)	0	0	2.60	−1.00	$-0.322^{+0.17}_{-0.25}$	$15.0^{+2.7}_{-3.4}$	$1.91^{+0.35}_{-0.33}$	14.1 ± 0.12	$-0.848^{+0.21}_{-0.2}$	5.63 ± 0.83
fgas−8σ(4)	0	0	2.60	−1.00	−0.28	14.0	1.92 ± 0.35	14.1 ± 0.12	$-0.893^{+0.16}_{-0.19}$	5.57 ± 0.56
fgas−8σ(3)	0	0	2.60	−1.00	−0.28	14.0	1.60	14.0 ± 0.14	$-0.885^{+0.17}_{-0.18}$	5.40 ± 0.52
M*−σ(10)	$1.70^{+2.0}_{-2.3}$	$-0.901^{+1.1}_{-0.94}$	$2.58^{+1.5}_{-0.74}$	-1.06 ± 0.16	$-0.232^{+0.12}_{-0.22}$	$13.6^{+3.5}_{-3.3}$	$1.89^{+0.47}_{-0.44}$	13.9 ± 0.22	$-1.28^{+0.43}_{-0.4}$	$5.48^{+2.2}_{-1.6}$
M*−σ(8)	0	0	$2.75^{+1.4}_{-0.87}$	-1.04 ± 0.15	$-0.263^{+0.14}_{-0.29}$	$14.4^{+3.2}_{-3.4}$	$1.85^{+0.48}_{-0.42}$	14.0 ± 0.18	$-1.15^{+0.4}_{-0.44}$	$4.73^{+1.6}_{-1.1}$
M*−σ(6)	0	0	2.60	−1.00	$-0.296^{+0.15}_{-0.25}$	$15.0^{+2.9}_{-3.7}$	$1.90^{+0.46}_{-0.39}$	14.0 ± 0.16	$-1.09^{+0.35}_{-0.42}$	$4.94^{+0.81}_{-0.92}$
M*−σ(4)	0	0	2.60	−1.00	−0.28	14.0	$1.92^{+0.42}_{-0.36}$	14.0 ± 0.16	$-1.20^{+0.32}_{-0.36}$	4.85 ± 0.57
M*−σ(3)	0	0	2.60	−1.00	−0.28	14.0	1.60	14.0 ± 0.21	$-1.17^{+0.33}_{-0.37}$	4.69 ± 0.55
M*−σ_fgas−4σ(10)	$1.60^{+2.0}_{-2.2}$	$-0.863^{+1.0}_{-0.9}$	$2.61^{+1.3}_{-0.73}$	-1.02 ± 0.13	$-0.268^{+0.15}_{-0.34}$	13.7 ± 3.3	$1.92^{+0.41}_{-0.36}$	14.0 ± 0.19	$-0.906^{+0.29}_{-0.31}$	$5.64^{+2.2}_{-1.5}$
M*−σ_fgas−4σ(8)	0	0	$2.89^{+1.3}_{-0.86}$	-1.01 ± 0.12	$-0.329^{+0.19}_{-0.41}$	$14.4^{+3.0}_{-3.3}$	$1.91^{+0.39}_{-0.35}$	14.0 ± 0.14	-0.814 ± 0.27	$4.81^{+1.4}_{-0.97}$
M*−σ_fgas−4σ(6)	0	0	2.60	−1.00	$-0.33^{+0.17}_{-0.27}$	$15.1^{+2.6}_{-3.4}$	$1.88^{+0.38}_{-0.34}$	14.0 ± 0.13	-0.809 ± 0.23	$5.28^{+0.72}_{-0.81}$
M*−σ_fgas−4σ(4)	0	0	2.60	−1.00	−0.28	14.0	$1.91^{+0.36}_{-0.3}$	14.0 ± 0.12	-0.867 ± 0.2	5.16 ± 0.51
M*−σ_fgas−4σ(3)	0	0	2.60	−1.00	−0.28	14.0	1.60	14.0 ± 0.15	-0.852 ± 0.2	5.05 ± 0.5

Table 4. Continued from previous page.

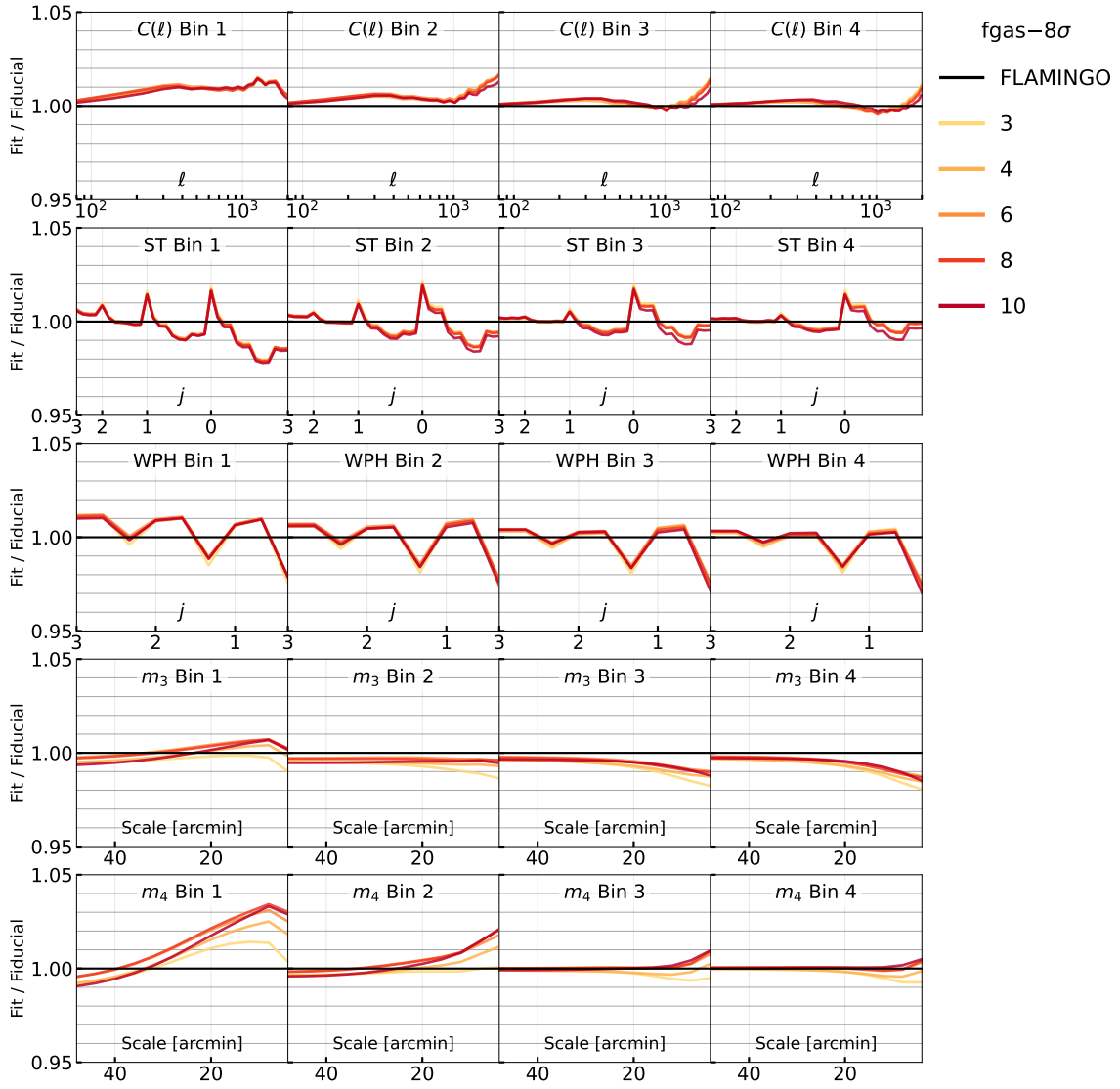

Figure 9. Similar to figure 5 but with BCM fitted to the strongest feedback variant in the FLAMINGO suite. The BCM error is still at the percent level. The elevated residual for low redshift m_4 is potentially related to both the baryon modelling error and/or m_4's sensitivity to noise.

E Baryonification error for strong and weak feedback variants

Following the discussion in section 6, we show here the residual errors for the best-fit BCM models applied to two extreme feedback variants from the FLAMINGO suite: the strongest (fgas-8σ) and the weakest (fgas$+2\sigma$). These figures are analogous to figure 5, which showed results for the fiducial case. For the strong feedback case (fgas-8σ), the BCM reproduces most statistics to within 2%, with a mild 3–4% excess in m_4 at low redshift and small angular scales (figure 9). This reflects both modeling limitations and the intrinsic noise in higher-order moments. For the weak feedback case (fgas$+2\sigma$), residuals remain within 2% across all statistics and scales (figure 10).

JCAP09(2025)073

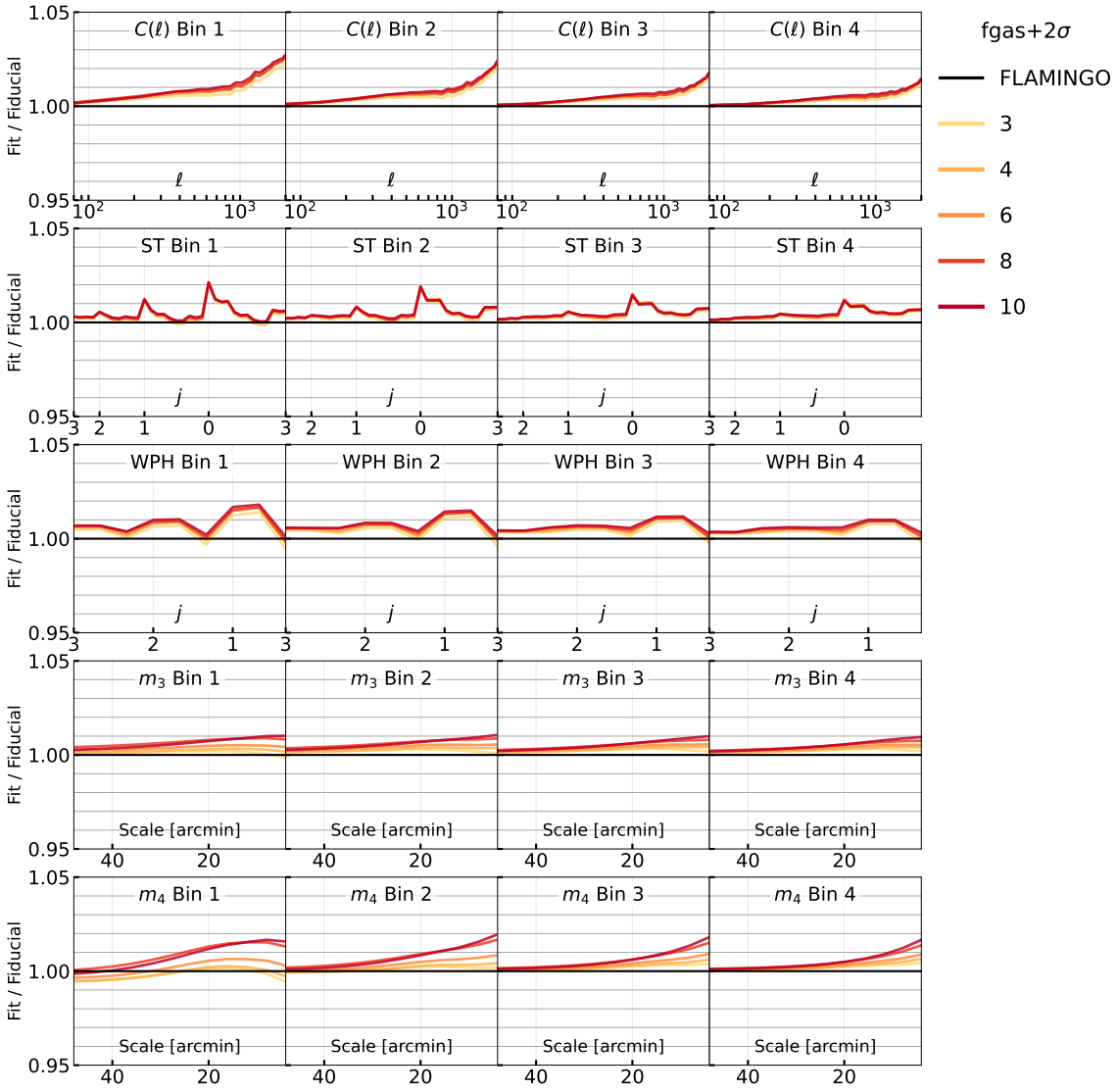

Figure 10. Similar to figure 5 but with BCM fitted to the weakest feedback variant in the FLAMINGO suite. The BCM error is still at the percent level.

F Cosmology dependence of baryonic suppression

To test how baryonic suppression of weak lensing summary statistics varies with cosmology, we compute the statistics for both the fiducial and LS8 cosmologies, and for each with the fiducial and strong-AGN (fgas−8σ) feedback variants. Figure 11 shows the relative difference in suppression between LS8 and fiducial cosmologies in the first redshift bin, which is sub-percent for all statistics and scales.

JCAP09(2025)073

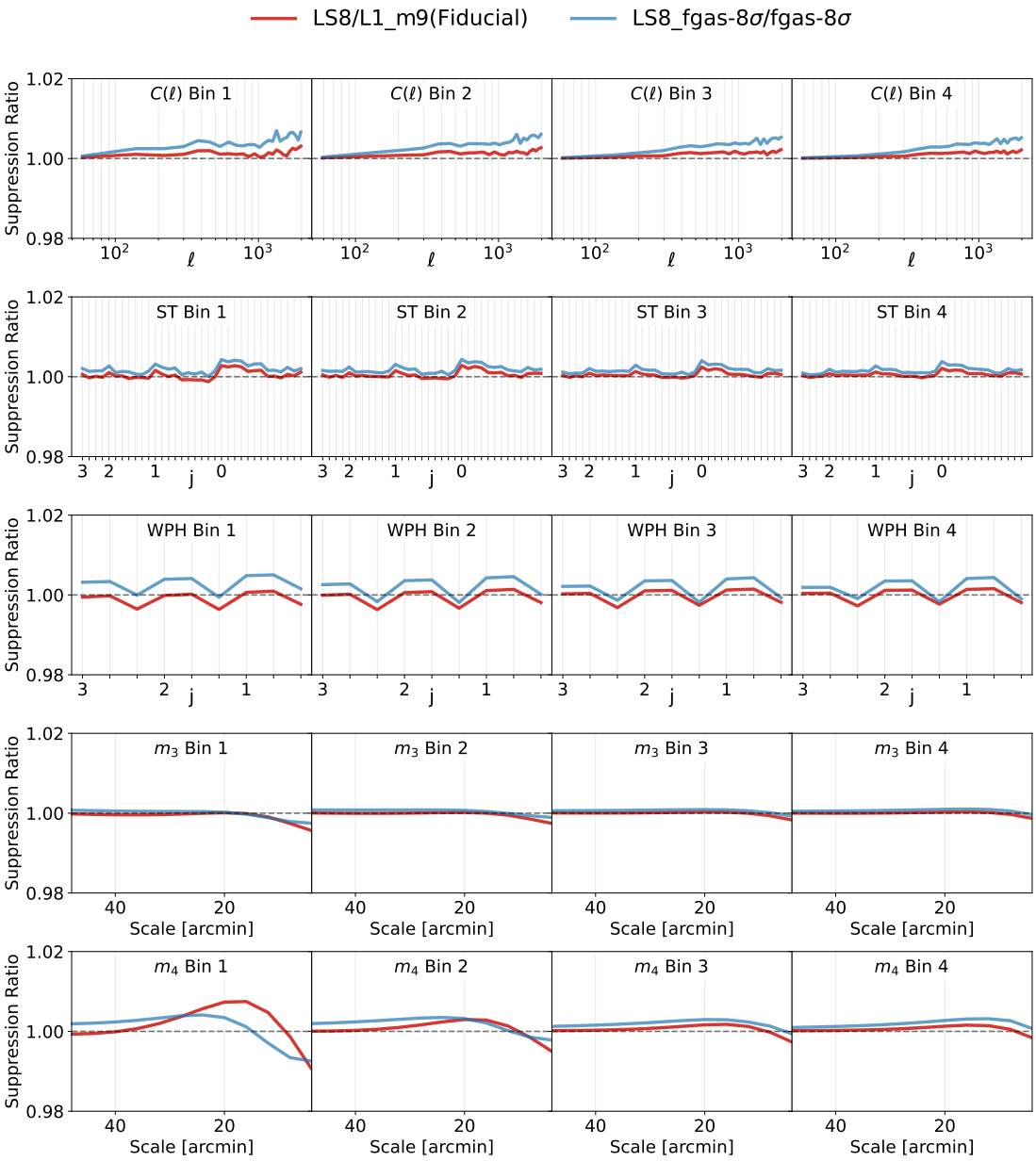

Figure 11. The sensitivity of the baryonic effect to cosmology for the first redshift bin. The y-axis shows the fractional change between the baryonic suppression of summary statistics at the LS8 cosmology compared to the suppression at the fiducial cosmology, for two different subgrid models: the fiducial models (LS8 and L1_m9) and the strong AGN feedback variants (LS8_fgas-8σ and fgas-8σ). Across both feedback variants, the amount of baryonic suppression changes only at the sub-percent level, indicating weak dependence on σ_8.

JCAP09(2025)073

References

[1] DES collaboration, *Dark Energy Survey year 3 results: cosmology from cosmic shear and robustness to data calibration*, *Phys. Rev. D* **105** (2022) 023514 [arXiv:2105.13543] [INSPIRE].

[2] DES collaboration, *Dark Energy Survey year 3 results: cosmology from cosmic shear and robustness to modeling uncertainty*, *Phys. Rev. D* **105** (2022) 023515 [arXiv:2105.13544] [INSPIRE].

[3] X. Li et al., *Hyper Suprime-Cam year 3 results: cosmology from cosmic shear two-point correlation functions*, *Phys. Rev. D* **108** (2023) 123518 [arXiv:2304.00702] [INSPIRE].

[4] R. Dalal et al., *Hyper Suprime-Cam year 3 results: cosmology from cosmic shear power spectra*, *Phys. Rev. D* **108** (2023) 123519 [arXiv:2304.00701] [INSPIRE].

[5] Kilo-Degree Survey and DES collaborations, *DES Y3 + KiDS-1000: consistent cosmology combining cosmic shear surveys*, *Open J. Astrophys.* **6** (2023) 2305.17173 [arXiv:2305.17173] [INSPIRE].

[6] D. Anbajagane et al., *The DECADE cosmic shear project IV: cosmological constraints from 107 million galaxies across 5,400 deg^2 of the sky*, arXiv:2502.17677 [INSPIRE].

[7] A.H. Wright et al., *KiDS-Legacy: cosmological constraints from cosmic shear with the complete Kilo-Degree Survey*, arXiv:2503.19441 [INSPIRE].

[8] J. Alsing, A.F. Heavens and A.H. Jaffe, *Cosmological parameters, shear maps and power spectra from CFHTLenS using Bayesian hierarchical inference*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 3272 [arXiv:1607.00008] [INSPIRE].

[9] A. Loureiro et al., *Almanac: weak lensing power spectra and map inference on the masked sphere*, *Open J. Astrophys.* **6** (2022) 2023 [arXiv:2210.13260] [INSPIRE].

[10] S.S. Boruah and E. Rozo, *Map-based cosmology inference with weak lensing — information content and its dependence on the parameter space*, *Mon. Not. Roy. Astron. Soc.* **527** (2023) L162 [arXiv:2307.00070] [INSPIRE].

[11] A.J. Zhou, X. Li, S. Dodelson and R. Mandelbaum, *Accurate field-level weak lensing inference for precision cosmology*, *Phys. Rev. D* **110** (2024) 023539 [arXiv:2312.08934] [INSPIRE].

[12] B. Dai and U. Seljak, *Multiscale flow for robust and optimal cosmological analysis*, *Proc. Nat. Acad. Sci.* **121** (2024) e2309624121 [arXiv:2306.04689] [INSPIRE].

[13] A.J. Zhou et al., *A Hamiltonian, post-Born, three-dimensional, on-the-fly ray tracing algorithm for gravitational lensing*, *JCAP* **10** (2024) 069 [arXiv:2405.12913] [INSPIRE].

[14] J. Fluri et al., *Cosmological constraints with deep learning from KiDS-450 weak lensing maps*, *Phys. Rev. D* **100** (2019) 063514 [arXiv:1906.03156] [INSPIRE].

[15] DES collaboration, *Dark Energy Survey year 3 results: cosmology with moments of weak lensing mass maps*, *Phys. Rev. D* **106** (2022) 083509 [arXiv:2110.10141] [INSPIRE].

[16] DES collaboration, *Dark Energy Survey year 3 results: likelihood-free, simulation-based wCDM inference with neural compression of weak-lensing map statistics*, *Mon. Not. Roy. Astron. Soc.* **536** (2024) 1303 [arXiv:2403.02314] [INSPIRE].

[17] G.A. Marques et al., *Cosmology from weak lensing peaks and minima with Subaru hyper Suprime-Cam survey first-year data*, *Mon. Not. Roy. Astron. Soc.* **528** (2024) 4513 [arXiv:2308.10866] [INSPIRE].

JCAP09(2025)073

[18] C.P. Novaes et al., *Cosmology from HSC Y1 weak lensing data with combined higher-order statistics and simulation-based inference*, *Phys. Rev. D* **111** (2025) 083510 [arXiv:2409.01301] [INSPIRE].

[19] DES collaboration, *Dark Energy Survey year 3 results: simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps. II. Cosmological results*, *Phys. Rev. D* **111** (2025) 063504 [arXiv:2405.10881] [INSPIRE].

[20] S. Cheng et al., *Cosmological constraints from weak lensing scattering transform using HSC Y1 data*, *JCAP* **01** (2025) 006 [arXiv:2404.16085] [INSPIRE].

[21] D. Anbajagane, C. Chang, H. Lee and M. Gatti, *Primordial non-Gaussianities with weak lensing: information on non-linear scales in the Ulagam full-sky simulations*, *JCAP* **03** (2024) 062 [arXiv:2310.02349] [INSPIRE].

[22] X. Liu et al., *Constraining $f(R)$ gravity theory using weak lensing peak statistics from the Canada-France-Hawaii-telescope lensing survey*, *Phys. Rev. Lett.* **117** (2016) 051101 [arXiv:1607.00184] [INSPIRE].

[23] C.T. Davies et al., *Constraining modified gravity with weak-lensing peaks*, *Mon. Not. Roy. Astron. Soc.* **533** (2024) 3546 [arXiv:2406.11958] [INSPIRE].

[24] O.Y. Gnedin, A.V. Kravtsov, A.A. Klypin and D. Nagai, *Response of dark matter halos to condensation of baryons: cosmological simulations and improved adiabatic contraction model*, *Astrophys. J.* **616** (2004) 16 [astro-ph/0406247] [INSPIRE].

[25] A.R. Duffy et al., *Impact of baryon physics on dark matter structures: a detailed simulation study of halo density profiles*, *Mon. Not. Roy. Astron. Soc.* **405** (2010) 2161 [arXiv:1001.3447] [INSPIRE].

[26] M.P. van Daalen, J. Schaye, C.M. Booth and C.D. Vecchia, *The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology*, *Mon. Not. Roy. Astron. Soc.* **415** (2011) 3649 [arXiv:1104.1174] [INSPIRE].

[27] E. Semboloni, H. Hoekstra and J. Schaye, *Effect of baryonic feedback on two- and three-point shear statistics: prospects for detection and improved modelling*, *Mon. Not. Roy. Astron. Soc.* **434** (2013) 148 [arXiv:1210.7303] [INSPIRE].

[28] D. Anbajagane, A.E. Evrard and A. Farahi, *Baryonic imprints on DM haloes: population statistics from dwarf galaxies to galaxy clusters*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 3441 [arXiv:2109.02713] [INSPIRE].

[29] M. and (Jon) Shao, D. Anbajagane and C. Chang, *Baryonic imprints on DM haloes: the concentration-mass relation in the C amels simulations*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 3258 [arXiv:2212.05964] [INSPIRE].

[30] I.G. McCarthy et al., *The FLAMINGO project: revisiting the S8 tension and the role of baryonic physics*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 5494 [arXiv:2309.07959] [INSPIRE].

[31] J.C. Broxterman et al., *The FLAMINGO project: baryonic impact on weak gravitational lensing convergence peak counts*, *Mon. Not. Roy. Astron. Soc.* **529** (2024) 2309 [arXiv:2312.08450] [INSPIRE].

[32] M. Schaller et al., *The FLAMINGO project: baryon effects on the matter power spectrum*, *Mon. Not. Roy. Astron. Soc.* **539** (2025) 1337 [arXiv:2410.17109] [INSPIRE].

[33] M. Schaller and J. Schaye, *An analytic redshift-independent formulation of baryonic effects on the matter power spectrum*, arXiv:2504.15633 [INSPIRE].

JCAP09(2025)073

[34] E. Semboloni et al., *Quantifying the effect of baryon physics on weak lensing tomography*, *Mon. Not. Roy. Astron. Soc.* **417** (2011) 2020 [arXiv:1105.1075] [INSPIRE].

[35] M. Yoon et al., *Constraints on cosmology and baryonic feedback with the deep lens survey using galaxy-galaxy and galaxy-mass power spectra*, *Astrophys. J.* **870** (2019) 111 [arXiv:1807.09195] [INSPIRE].

[36] A. Amon and G. Efstathiou, *A non-linear solution to the S_8 tension?*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 5355 [arXiv:2206.11794] [INSPIRE].

[37] DES collaboration, *Beyond the 3rd moment: a practical study of using lensing convergence CDFs for cosmology with DES Y3*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 5530 [arXiv:2308.03863] [INSPIRE].

[38] G. Piccirilli, M. Zennaro, C. García-García and D. Alonso, *Robust cosmic shear with small-scale nulling*, arXiv:2502.17339 [INSPIRE].

[39] O. Truttero, J. Zuntz, A. Pourtsidou and N. Robertson, *Baryon-free S_8 tension with stage IV cosmic shear surveys*, arXiv:2410.18191 [DOI:10.33232/001c.129965] [INSPIRE].

[40] M. Vogelsberger, F. Marinacci, P. Torrey and E. Puchwein, *Cosmological simulations of galaxy formation*, *Nature Rev. Phys.* **2** (2020) 42 [arXiv:1909.07976] [INSPIRE].

[41] M. Velliscig et al., *The impact of galaxy formation on the total mass, mass profile and abundance of haloes*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 2641 [arXiv:1402.4461] [INSPIRE].

[42] M.P. van Daalen and J. Schaye, *The contributions of matter inside and outside of haloes to the matter power spectrum*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 2247 [arXiv:1501.05950] [INSPIRE].

[43] S.N.B. Debackere, J. Schaye and H. Hoekstra, *The impact of the observed baryon distribution in haloes on the total matter power spectrum*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 2285 [arXiv:1908.05765] [INSPIRE].

[44] D. Anbajagane et al., *Stellar property statistics of massive haloes from cosmological hydrodynamics simulations: common kernel shapes*, *Mon. Not. Roy. Astron. Soc.* **495** (2020) 686 [arXiv:2001.02283] [INSPIRE].

[45] S.H. Lim et al., *Properties of the ionized CGM and IGM: tests for galaxy formation models from the Sunyaev-Zel'dovich effect*, *Mon. Not. Roy. Astron. Soc.* **504** (2021) 5131 [arXiv:2007.11583].

[46] E. Lee et al., *A multisimulation study of relativistic SZ temperature scalings in galaxy clusters and groups*, *Mon. Not. Roy. Astron. Soc.* **517** (2022) 5303 [arXiv:2207.05834] [INSPIRE].

[47] R. Stiskalek, D.J. Bartlett, H. Desmond and D. Anbajagane, *The scatter in the galaxy-halo connection: a machine learning analysis*, *Mon. Not. Roy. Astron. Soc.* **514** (2022) 4026 [arXiv:2202.14006] [INSPIRE].

[48] D. Anbajagane et al., *Galaxy velocity bias in cosmological simulations: towards per cent-level calibration*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 2980 [arXiv:2110.01683] [INSPIRE].

[49] M. Gebhardt et al., *Cosmological baryon spread and impact on matter clustering in CAMELS*, *Mon. Not. Roy. Astron. Soc.* **529** (2024) 4896 [arXiv:2307.11832] [INSPIRE].

[50] N.E. Chisari et al., *The impact of baryons on the matter power spectrum from the Horizon-AGN cosmological hydrodynamical simulation*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 3962 [arXiv:1801.08559] [INSPIRE].

JCAP09(2025)073

[51] M.P. van Daalen, I.G. McCarthy and J. Schaye, *Exploring the effects of galaxy formation on matter clustering through a library of simulation power spectra*, *Mon. Not. Roy. Astron. Soc.* **491** (2020) 2424 [arXiv:1906.00968] [INSPIRE].

[52] CAMELS collaboration, *The CAMELS project: Cosmology and Astrophysics with MachinE Learning Simulations*, *Astrophys. J.* **915** (2021) 71 [arXiv:2010.00619] [INSPIRE].

[53] A.J. Mead et al., *A hydrodynamical halo model for weak-lensing cross correlations*, *Astron. Astrophys.* **641** (2020) A130 [arXiv:2005.00009] [INSPIRE].

[54] I.G. McCarthy, J. Schaye, S. Bird and A.M.C. Le Brun, *The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 2936 [arXiv:1603.02702] [INSPIRE].

[55] A. Schneider and R. Teyssier, *A new method to quantify the effects of baryons on the matter power spectrum*, *JCAP* **12** (2015) 049 [arXiv:1510.06034] [INSPIRE].

[56] A. Schneider et al., *Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation*, *JCAP* **03** (2019) 020 [arXiv:1810.08629] [INSPIRE].

[57] D. Anbajagane, S. Pandey and C. Chang, *Map-level baryonification: efficient modelling of higher-order correlations in the weak lensing and thermal Sunyaev-Zeldovich fields*, *Open J. Astrophys.* **7** (2024) [arXiv:2409.03822] [INSPIRE].

[58] G. Aricò et al., *Modelling the large-scale mass density field of the universe as a function of cosmology and baryonic physics*, *Mon. Not. Roy. Astron. Soc.* **495** (2020) 4800 [arXiv:1911.08471] [INSPIRE].

[59] S.K. Giri and A. Schneider, *Emulation of baryonic effects on the matter power spectrum and constraints from galaxy cluster data*, *JCAP* **12** (2021) 046 [arXiv:2108.08863] [INSPIRE].

[60] DES collaboration, *Constraining the baryonic feedback with cosmic shear using the DES year-3 small-scale measurements*, *Mon. Not. Roy. Astron. Soc.* **518** (2023) 5340 [arXiv:2206.08591] [INSPIRE].

[61] G. Aricò et al., *DES Y3 cosmic shear down to small scales: constraints on cosmology and baryons*, *Astron. Astrophys.* **678** (2023) A109 [arXiv:2303.05537] [INSPIRE].

[62] DES collaboration, *Weak lensing combined with the kinetic Sunyaev–Zel'dovich effect: a study of baryonic feedback*, *Mon. Not. Roy. Astron. Soc.* **534** (2024) 655 [arXiv:2404.06098] [INSPIRE].

[63] G. Aricò et al., *Simultaneous modelling of matter power spectrum and bispectrum in the presence of baryons*, *Mon. Not. Roy. Astron. Soc.* **503** (2021) 3596 [arXiv:2009.14225] [INSPIRE].

[64] M.E. Lee et al., *Comparing weak lensing peak counts in baryonic correction models to hydrodynamical simulations*, *Mon. Not. Roy. Astron. Soc.* **519** (2023) 573 [arXiv:2201.08320] [INSPIRE].

[65] J. Schaye et al., *The FLAMINGO project: cosmological hydrodynamical simulations for large-scale structure and galaxy cluster surveys*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 4978 [arXiv:2306.04024] [INSPIRE].

[66] R. Kugel et al., *FLAMINGO: calibrating large cosmological hydrodynamical simulations with machine learning*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 6103 [arXiv:2306.05492] [INSPIRE].

JCAP09(2025)073

[67] S. Cheng, Y.-S. Ting, B. Ménard and J. Bruna, *A new approach to observational cosmology using the scattering transform*, *Mon. Not. Roy. Astron. Soc.* **499** (2020) 5902 [arXiv:2006.08561] [INSPIRE].

[68] E. Allys et al., *New interpretable statistics for large scale structure analysis and generation*, *Phys. Rev. D* **102** (2020) 103506 [arXiv:2006.06298] [INSPIRE].

[69] S. Pandey et al., *Accurate connected modeling of gas thermodynamics and matter distribution*, *Phys. Rev. D* **111** (2025) 043529 [arXiv:2401.18072] [INSPIRE].

[70] J.F. Navarro, C.S. Frenk and S.D.M. White, *The structure of cold dark matter halos*, *Astrophys. J.* **462** (1996) 563 [astro-ph/9508025] [INSPIRE].

[71] J.F. Navarro, C.S. Frenk and S.D.M. White, *A universal density profile from hierarchical clustering*, *Astrophys. J.* **490** (1997) 493 [astro-ph/9611107] [INSPIRE].

[72] M. Oguri and M. Takada, *Combining cluster observables and stacked weak lensing to probe dark energy: self-calibration of systematic uncertainties*, *Phys. Rev. D* **83** (2011) 023008 [arXiv:1010.0744] [INSPIRE].

[73] B. Diemer and A.V. Kravtsov, *A universal model for halo concentrations*, *Astrophys. J.* **799** (2015) 108 [arXiv:1407.4730] [INSPIRE].

[74] C.-H. To et al., *Deciphering baryonic feedback with galaxy clusters*, *JCAP* **07** (2024) 037 [arXiv:2402.00110] [INSPIRE].

[75] DES collaboration, *Weak lensing combined with the kinetic Sunyaev–Zel'dovich effect: a study of baryonic feedback*, *Mon. Not. Roy. Astron. Soc.* **534** (2024) 655 [arXiv:2404.06098] [INSPIRE].

[76] J. Fluri et al., *Full wCDM analysis of KiDS-1000 weak lensing maps using deep learning*, *Phys. Rev. D* **105** (2022) 083518 [arXiv:2201.07771] [INSPIRE].

[77] J. Borrow, M. Schaller, R.G. Bower and J. Schaye, *Sphenix: smoothed particle hydrodynamics for the next generation of galaxy formation simulations*, *Mon. Not. Roy. Astron. Soc.* **511** (2021) 2367 [arXiv:2012.03974].

[78] SWIFT collaboration, *Swift: a modern highly parallel gravity and smoothed particle hydrodynamics solver for astrophysical and cosmological applications*, *Mon. Not. Roy. Astron. Soc.* **530** (2024) 2378 [arXiv:2305.13380] [INSPIRE].

[79] S. Ploeckinger and J. Schaye, *Radiative cooling rates, ion fractions, molecule abundances, and line emissivities including self-shielding and both local and metagalactic radiation fields*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) 4857 [arXiv:2006.14322].

[80] J. Schaye and C. Dalla Vecchia, *On the relation between the Schmidt and Kennicutt-Schmidt star formation laws and its implications for numerical simulations*, *Mon. Not. Roy. Astron. Soc.* **383** (2008) 1210 [arXiv:0709.0292] [INSPIRE].

[81] R.P.C. Wiersma et al., *Chemical enrichment in cosmological, smoothed particle hydrodynamics simulations*, *Mon. Not. Roy. Astron. Soc.* **399** (2009) 574 [arXiv:0902.1535] [INSPIRE].

[82] W. Elbers et al., *An optimal non-linear method for simulating relic neutrinos*, *Mon. Not. Roy. Astron. Soc.* **507** (2021) 2614 [arXiv:2010.07321] [INSPIRE].

[83] C. Dalla Vecchia and J. Schaye, *Simulating galactic outflows with kinetic supernova feedback*, *Mon. Not. Roy. Astron. Soc.* **387** (2008) 1431 [arXiv:0801.2770] [INSPIRE].

[84] E. Chaikin et al., *A thermal-kinetic subgrid model for supernova feedback in simulations of galaxy formation*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 3709 [arXiv:2211.04619].

[85] C.M. Booth, J. Schaye, S. Heinz and E. Wilcots, *Simulations of the growth of black holes and feedback from active galactic nuclei*, *AIP Conf. Proc.* **1201** (2009) 21.

[86] F. Huško et al., *Spin-driven jet feedback in idealized simulations of galaxy groups and clusters*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 3750 [arXiv:2206.06402] [INSPIRE].

[87] DES collaboration, *Dark Energy Survey year 3 results: cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev. D* **105** (2022) 023520 [arXiv:2105.13549] [INSPIRE].

[88] A. Amon et al., *Consistent lensing and clustering in a low-S8 universe with BOSS, DES year 3, HSC year 1, and KiDS-1000*, *Mon. Not. Roy. Astron. Soc.* **518** (2023) 477 [arXiv:2202.07440] [INSPIRE].

[89] K.M. Górski et al., *HEALPix — a framework for high resolution discretization, and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [astro-ph/0409513] [INSPIRE].

[90] DES and EBOSS collaborations, *Dark Energy Survey year 3 results: clustering redshifts — calibration of the weak lensing source redshift distributions with redMaGiC and BOSS/eBOSS*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 1223 [arXiv:2012.08569] [INSPIRE].

[91] DES collaboration, *Dark Energy Survey year 3 results: redshift calibration of the weak lensing source galaxies*, *Mon. Not. Roy. Astron. Soc.* **505** (2021) 4249 [arXiv:2012.08566] [INSPIRE].

[92] A. Petri, Z. Haiman and M. May, *Validity of the Born approximation for beyond Gaussian weak lensing observables*, *Phys. Rev. D* **95** (2017) 123503 [arXiv:1612.00852] [INSPIRE].

[93] J.C. Broxterman et al., *The FLAMINGO project: cosmology with the redshift dependence of weak gravitational lensing peaks*, *Mon. Not. Roy. Astron. Soc.* **538** (2025) 755 [arXiv:2412.02736] [INSPIRE].

[94] EUCLID collaboration, *Euclid preparation: II. The EuclidEmulator — a tool to compute the cosmology dependence of the nonlinear matter power spectrum*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 5509 [arXiv:1809.04695] [INSPIRE].

[95] A. Tenneti, N.Y. Gnedin and Y. Feng, *Impact of baryonic physics on intrinsic alignments*, *Astrophys. J.* **834** (2017) 169 [arXiv:1607.07140] [INSPIRE].

[96] A. Soussana et al., *The impact of AGN feedback on galaxy intrinsic alignments in the horizon simulations*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 4268 [arXiv:1908.11665] [INSPIRE].

[97] DES collaboration, *Dark Energy Survey year 3 results: simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps. Validation on simulations*, *Phys. Rev. D* **109** (2024) 063534 [arXiv:2310.17557] [INSPIRE].

[98] S. Mallat, *Group invariant scattering*, *Commun. Pure Appl. Math.* **65** (2012) 1331.

[99] J. Carron, *On the incompleteness of the moment and correlation function hierarchy as probes of the lognormal field*, *Astrophys. J.* **738** (2011) 86 [arXiv:1105.4467] [INSPIRE].

[100] E. Allys et al., *The RWST, a comprehensive statistical description of the non-Gaussian structures in the ISM*, *Astron. Astrophys.* **629** (2019) A115 [arXiv:1905.01372] [INSPIRE].

[101] S. Mallat, S. Zhang and G. Rochette, *Phase harmonic correlations and convolutional neural networks*, arXiv:1810.12136.

[102] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1 [astro-ph/0112551] [INSPIRE].

JCAP09(2025)073

[103] H.-J. Huang, T. Eifler, R. Mandelbaum and S. Dodelson, *Modelling baryonic physics in future weak lensing surveys*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 1652 [arXiv:1809.01146] [INSPIRE].

[104] D. Grandón et al., *Impact of baryonic feedback on HSC-Y1 weak lensing non-Gaussian statistics*, *Phys. Rev. D* **110** (2024) 103539 [arXiv:2403.03807] [INSPIRE].

[105] J. Bradbury et al., *JAX: composable transformations of Python+NumPy programs*, http://github.com/google/jax (2018).

[106] E. Bingham et al., *Pyro: deep universal probabilistic programming*, arXiv:1810.09538.

[107] D. Phan, N. Pradhan and M. Jankowiak, *Composable effects for flexible and accelerated probabilistic programming in NumPyro*, arXiv:1912.11554 [INSPIRE].

[108] R.M. Neal, *Handbook of Markov chain Monte Carlo*, Chapman and Hall/CRC (2011) [DOI:10.1201/b10905] [arXiv:1206.1901] [INSPIRE].

[109] M. Betancourt, *A conceptual introduction to Hamiltonian Monte Carlo*, arXiv:1701.02434 [INSPIRE].

[110] W. Elbers et al., *The FLAMINGO project: the coupling between baryonic feedback and cosmology in light of the S8 tension*, *Mon. Not. Roy. Astron. Soc.* **537** (2025) 2160 [arXiv:2403.12967] [INSPIRE].

[111] M.M. Rau et al., *Weak lensing tomographic redshift distribution inference for the Hyper Suprime-Cam Subaru strategic program three-year shape catalogue*, *Mon. Not. Roy. Astron. Soc.* **524** (2023) 5109 [arXiv:2211.16516] [INSPIRE].

[112] D. Anbajagane et al., *The DECADE cosmic shear project II: photometric redshift calibration of the source galaxy sample*, arXiv:2502.17675 [INSPIRE].

[113] A.H. Wright et al., *KiDS-legacy: redshift distributions and their calibration*, arXiv:2503.19440 [INSPIRE].

[114] B. Jain and L.V. Van Waerbeke, *Statistics of dark matter halos from gravitational lensing*, *Astrophys. J. Lett.* **530** (2000) L1 [astro-ph/9910459] [INSPIRE].

[115] B. Xu, N. Wang, T. Chen and M. Li, *Empirical evaluation of rectified activations in convolutional network*, arXiv:1505.00853 [INSPIRE].

[116] N. Srivastava et al., *Dropout: a simple way to prevent neural networks from overfitting*, *J. Machine Learning Res.* **15** (2014) 1929 [INSPIRE].

[117] D.P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, arXiv:1412.6980 [INSPIRE].

JCAP09(2025)073