

The FLAMINGO project: baryonic impact on weak gravitational lensing convergence peak counts

Jeger C. Broxterman^{1,2}*, Matthieu Schaller^{1,2}, Joop Schaye^{1,2}, Henk Hoekstra^{1,2}, Konrad Kuijken², John C. Helly³, Roi Kugel^{1,2}, Joey Braspenning², Willem Elbers^{1,3}, Carlos S. Frenk³, Juliana Kwan⁴, Ian G. McCarthy^{1,4}, Jaime Salcido^{1,4}, Marcel P. van Daalen^{1,2} and Bert Vandenbroucke^{1,2}

¹Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, NL-2300 RA Leiden, The Netherlands

²Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, The Netherlands

³Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

⁴Astrophysics Research Institute, Liverpool John Moores University, Liverpool L3 5RF, UK

Accepted 2024 March 5. Received 2024 February 19; in original form 2023 December 13

ABSTRACT

Weak gravitational lensing convergence peaks, the local maxima in weak lensing convergence maps, have been shown to contain valuable cosmological information complementary to commonly used two-point statistics. To exploit the full power of weak lensing for cosmology, we must model baryonic feedback processes because these reshape the matter distribution on non-linear and mildly non-linear scales. We study the impact of baryonic physics on the number density of weak lensing peaks using the FLAMINGO cosmological hydrodynamical simulation suite. We generate ray-traced full-sky convergence maps mimicking the characteristics of a Stage IV weak lensing survey. We compare the number densities of peaks in simulations that have been calibrated to reproduce the observed galaxy mass function and cluster gas fraction or to match a shifted version of these, and that use either thermally driven or jet active galactic nucleus feedback. We show that the differences induced by realistic baryonic feedback prescriptions (typically 5–30 per cent for $\kappa = 0.1$ –0.4) are smaller than those induced by reasonable variations in cosmological parameters (20–60 per cent for $\kappa = 0.1$ –0.4) but must be modelled carefully to obtain unbiased results. The reasons behind these differences can be understood by considering the impact of feedback on halo masses, or by considering the impact of different cosmological parameters on the halo mass function. Our analysis demonstrates that, for the range of models we investigated, the baryonic suppression is insensitive to changes in cosmology up to $\kappa \approx 0.4$ and that the higher κ regime is dominated by Poisson noise and cosmic variance.

Key words: gravitational lensing; weak – methods: numerical – large-scale structure of Universe – cosmology: theory.

1. INTRODUCTION

Over the last few decades, the spatially flat Lambda cold dark matter (Λ CDM) model has become the generally accepted standard cosmological model. While it depends on only six free parameters, it can predict several key observations with great accuracy, including fluctuations of the cosmic microwave background (CMB, Planck Collaboration VI 2020), galaxy clustering (Anderson et al. 2014), and Type Ia supernovae (Abbott et al. 2019) (for a recent review see Lahav & Liddle 2022). As the observations become increasingly more constraining, tensions between different cosmological probes have emerged, most notably on the value of the Hubble constant, H_0 , and S_8 (e.g. Hildebrandt et al. 2017; Abdalla et al. 2022; Schöneberg et al. 2022; Clark et al. 2023). Understanding the origin of these tensions, which may lead to new physics beyond the Λ CDM model, is one of the major goals of modern cosmology.

One of the key tools to constrain cosmology is cosmic shear, the slight distortion of distant galaxy images through weak gravitational lensing (weak lensing or WL) by the large-scale structure (LSS) of the Universe. It allows us to estimate the projected matter distribution along the line of sight, which can be related to the underlying cosmology. For reviews, see Bartelmann & Schneider (2001), Hoekstra & Jain (2008), and Kilbinger (2015). Upcoming next-generation weak gravitational lensing surveys, carried out from space by the *Roman* (Spergel et al. 2015) and the recently-launched *Euclid* (Laureijs et al. 2011) satellites, and from the ground by Rubin (LSST Science Collaboration 2009), collectively referred to as Stage IV surveys, will cover almost the entirety of the observable sky suitable for WL. They will reach unprecedented depths as well as measure the WL signal as a function of redshift, allowing them to quantify the evolution of the matter distribution in the Universe. Jointly, these missions aim to provide insight into the nature of dark matter (DM), dark energy, and the expansion history of the Universe.

Typically, WL surveys use two-point statistics, either in configuration space or harmonic space, to constrain the cosmological model (e.g. Hamana et al. 2020; Asgari et al. 2021; Amon et al. 2022).

* E-mail: broxterman@strw.leidenuniv.nl

Astrophysical feedback processes [e.g. supernova (SN) explosions and active galactic nuclei (AGNs)] reshape the matter distribution on partly the same scales that are typically probed by the two-point inferences, thereby complicating the analysis. It has been shown that this so-called baryonic feedback suppresses the matter power spectrum on scales of $k \lesssim 0.1 \, h \text{Mpc}^{-1}$ and enhances the power on even smaller scales (e.g. van Daalen et al. 2011; Chisari et al. 2018; Schneider et al. 2020; van Daalen, McCarthy & Schaye 2020; Salcido et al. 2023; Schaye et al. 2023). When not taking these baryonic effects into account, strong biases in inferences may arise for WL statistics (e.g. Semboloni et al. 2011; Semboloni, Hoekstra & Schaye 2013; Gouin et al. 2019; Weiss et al. 2019).

Two-point statistics encapsulate all cosmological information in the underlying field if it can be described as a Gaussian random field. While today this applies to the largest scales, the cosmological information on smaller scales, which correspond to the regime of non-linear collapse and contain additional information, is not fully captured by Gaussian statistics. Over the last decade, interest has grown in non-Gaussian statistics, which are able to probe this regime. Examples of commonly used beyond-Gaussian statistics are the bispectrum (e.g. Dodelson & Zhang 2005), Minkowski functionals (e.g. Kratochvil et al. 2012), higher order moments of the convergence field (e.g. Petri et al. 2013), WL peaks or voids (e.g. Kratochvil, Haiman & May 2010; Davies et al. 2021, 2022), and Betti numbers (e.g. Feldbrugge et al. 2019). The addition of non-Gaussian statistics can provide tighter cosmological constraints (e.g. Euclid Collaboration 2023) and help discriminate between cosmological and baryonic effects (e.g. Semboloni et al. 2013). In general, the baryonic impact on these statistics is not understood as well as it is for the two-point statistics and only approximate methods, e.g. by using a halo model (e.g. Sabyr et al. 2022; Akgari, Mead & Heymans 2023), for calculating these quantities exist.

In this paper, we choose to focus on one of these non-Gaussian statistics, namely WL peaks, which correspond to local maxima in the WL convergence field. WL peaks have been found to be highly complementary to typical two-point statistics, both in simulations (e.g. Dietrich & Hartlap 2010) and observations (e.g. Marques et al. 2024). Whereas peaks can arise due to chance alignments of haloes along the line of sight, the highest peaks primarily stem from a single high-mass halo along the line of sight (Li et al. 2019). Lower peaks are typically caused by multiple smaller haloes aligned along the line of sight, but dominate the cosmological information contained in the peaks (Yang et al. 2011). Peak counts are thus directly sensitive to the number of haloes and therefore directly probe the halo mass function (HMF), which depends strongly on cosmology (e.g. Kaiser 1986; Tinker et al. 2008; McClintock et al. 2019), such that changes in cosmology directly influence the number density of WL peaks. Compared with other probes of high-density regions, such as cluster X-ray luminosity and temperature, or cluster optical richness, WL peak counts are a more direct tracer of the total mass in haloes, and they are not plagued by uncertainties arising from a set of assumptions regarding the dynamical state of the galaxy clusters (e.g. hydrostatic equilibrium or relaxedness), nor do they require scaling relations between mass and luminosity tracers. Therefore, peaks are an ideal probe to enhance WL inferences in constraining cosmology.

There are two main approaches to studying the impact of baryons on WL statistics. The first approach, which we adopt, uses full-hydrodynamical simulations employing subgrid models of relevant baryonic processes. Despite being computationally more expensive, this method has the important advantage of being fully self-consistent and allowing to compare the WL signal to non-gravitational probes

such as X-ray and the Sunyaev–Zel’dovich effect (SZE). The second is to use N -body simulations to model the evolution of DM, and to modify the shape of these dark matter only (DMO) simulations using a baryonic correction model (BCM) (e.g. Schneider et al. 2019; Aricò et al. 2020). Lee et al. (2023) compared an N -body + BCM and its corresponding hydrodynamical simulation and found that current BCM approaches that are calibrated on the power spectrum are not flexible enough for a WL peak count inference for Stage IV WL surveys, which is the focus of this paper. However, it is not a priori expected that a calibration on the power spectrum can recover all the peak properties as not all information in the peaks is captured by the power spectrum. Lu & Haiman (2021) use a similar but adapted BCM in the context of a Stage III inference and they found that the degeneracy between cosmological and baryonic parameters may be broken by considering peaks combined with the power spectrum.

Peak counts have been studied in the context of hydrodynamical simulations before. In general, it has been found that peak counts are sensitive to the baryonic contribution, and hence, for cosmological inferences based on peak counts, baryons have to be considered (e.g. Yang et al. 2013; Coulton et al. 2020). Similarly, the impact of neutrinos on peak counts has been studied using the BAHAMAS simulations (McCarthy et al. 2017, 2018). Fong et al. (2019) found that depending on the neutrino mass, either baryonic or neutrino effects dominate, and both effects should be accounted for in a proper WL peak analysis. Recently, Ferlito et al. (2023) carried out a comparison of WL peaks in cosmological hydrodynamical simulations, primarily focussing on WL convergence maps constructed using the MilleniumTNG (MTNG) simulation suite (Pakmor et al. 2023). Their analysis focuses on the contribution of neutrinos to WL peak counts. While we also look at the impact of massive neutrinos, we concentrate on the impact of baryonic feedback processes on WL peaks. Ferlito et al. (2023) also studied the baryonic impact on WL peaks by comparing their simulations with their corresponding DMO runs, as well as with simulations from the literature. As these simulations may differ significantly in terms of code, subgrid physics, cosmology, and resolution, a direct interpretation of the impact of the baryonic effect is not straightforward. Here, we instead use a consistent suite of simulations that systematically varies baryonic feedback strength.

In this paper, we explore the impact of baryonic physics on WL peak counts in the FLAMINGO simulation suite (Kugel et al. 2023; Schaye et al. 2023). The hydrodynamical simulations were calibrated, using machine-learning techniques, to reproduce the observed present-day gas fractions in clusters and the galaxy stellar mass function (SMF). The suite includes separately calibrated models that systematically vary these observables. In this way, we can directly relate changes in key observables, induced by feedback variations, to differences in WL peak counts. Additionally, the suite contains variations in cosmological parameters and neutrino masses, allowing us to compare the impact of cosmology with that of baryonic physics. To be able to quantify the changes induced by astrophysical feedback processes on next-generation WL surveys, we carry out a dedicated full-sky analysis at a high angular resolution in which the characteristics of a Stage IV WL survey are incorporated. The signal is determined for virtual observers which were placed within the simulation volume using a backward ray-tracing methodology and spherical harmonics on the full-sky sphere.

This paper is organized as follows. In Section 2, we introduce the relevant WL and ray-tracing theory. Section 3 introduces the FLAMINGO simulation suite, the key features of the different baryonic and cosmology models, and the construction of the WL maps. Here, we also validate our approach by quantifying the

sensitivity to different choices of smoothing, interpolation scheme, and angular resolution. In Section 4, we start by quantifying our level of numerical convergence in terms of box size and numerical resolution and the impact of cosmic variance on our signal, finding that the measured peak statistics are robust. Then, we compare simulation variations that were calibrated to different values of gas fractions in clusters and the galaxy SMF. In this way, we are able to understand the relation between stronger feedback, leading to lower gas fractions in clusters or lower galaxy SMFs, and the observed WL peak distribution. We then compare the different AGN subgrid models and the different cosmology variations in Section 5, where we also look at the separability of the baryonic and cosmological effects. We conclude by comparing the differences induced by baryonic feedback variations with those due to changes in cosmology. Our main results are summarized in Section 6.

2. WL THEORY

We will generate WL convergence signals from pixelized surface mass density maps discretized in redshift as seen by a virtual observer within a simulation. Before describing the construction of the convergence maps, we first summarize the main equations relevant to WL. For an extensive recent review, see Kilbinger (2015). We assume a flat Friedmann–Lemaître–Robertson–Walker (FLRW) metric such that the comoving angular diameter distance, $f_k(\chi)$, equals the comoving line of sight distance, χ . In WL, where deflections are small, the deflection field, α , which at each angular position θ describes the change in the position of a light ray perpendicular to the direction of travel, can be expressed in terms of the Newtonian gravitational potential Φ as

$$d\alpha = -\frac{2}{c^2} \nabla_{\perp} \Phi d\chi, \quad (1)$$

where c is the speed of light and the gradient, $\nabla_{\perp} = (\partial/\partial\beta_1, \partial/\partial\beta_2)$, is evaluated perpendicular to the light ray's direction of propagation. Integration over comoving distance yields the angular position of the light rays as seen by the observer:

$$\beta(\theta, \chi) = \beta(\theta, 0) - \frac{2}{c^2} \int_0^{\chi} d\chi' \frac{\chi - \chi'}{\chi \chi'} \nabla_{\perp} \Phi(\beta(\theta, \chi'), \chi'), \quad (2)$$

where $\beta(\theta, 0) = \theta$ is the observed angular position. The deflection of the photon path as it passes along a non-uniform matter distribution is described by the distortion matrix A , which is the derivative of the angular position and given by

$$A_{ij}(\theta, \chi) \equiv \frac{\partial \beta}{\partial \theta} = \delta_{ij} - \frac{2}{c^2} \int_0^{\chi} d\chi' \frac{\chi - \chi'}{\chi \chi'} \partial_i \partial_j \Phi(\beta(\theta, \chi'), \chi'), \quad (3)$$

where i and j are taken over the two angular coordinates on the sphere and δ_{ij} is the Kronecker delta function. Conventionally, the matrix is decomposed as

$$A \equiv \begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix} \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \approx \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 + \omega \\ -\gamma_2 - \omega & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (4)$$

which we refer to as the magnification matrix. Here, we assumed that the image rotation, described by the rotation angle ω , is small, as shown by Jain, Seljak & White (2000). κ is the lensing convergence, which we will use to quantify the WL strength and $\gamma = \gamma_1 + i\gamma_2$ is the lensing shear.

In WL, where the deflection angles are small, a common approach is to evaluate the deflection field on unperturbed photon paths, in which case the dependence of the angular position on comoving distance vanishes (i.e. $\beta(\theta, \chi) \cong \beta(\theta)$). This approximation is referred to as the Born approximation. The accuracy of the Born approximation has been well established for the reconstruction of the convergence angular power spectrum and cosmological parameter inferences based on it (e.g. Giocoli et al. 2016; Hilbert et al. 2020). For non-Gaussian statistics, the impact and validity of the Born approximation remain uncertain, as it has been found to impact galaxy–galaxy lensing shear profiles (Simon & Hilbert 2018), the CMB lensing bispectrum (Pratten & Lewis 2016), and higher order moments of convergence profiles, where Petri, Haiman & May (2017) find a 2.5σ bias on cosmological parameter estimations for a Rubin-like survey. Similarly, Lu & Haiman (2021) find that parameter inferences based on the Born approximation and peak counts can be up to 2σ biased for Hyper Suprime-Cam-like surveys.

As our analysis aims to mimic a *Euclid*-like signal, which will give tighter constraints, we expect this bias to be even more significant, stressing the need to adopt a beyond-Born approach. In our analysis, we will use the FLAMINGO mass shells. These shells are discrete full-sky maps spaced regularly in redshift ($\Delta z = 0.05$ between $z = 0$ and 3). To analyse the maps, the theoretical equations (2) and (3) need to be discretized. We choose to apply a backward ray-tracing method, as introduced by Jain et al. (2000). Here, starting from the position of a virtual observer, we track a ray's trajectory back in time as it is deflected by the gravitational potential of discrete matter shells along its propagation. For a detailed description of the deflection of the rays using discrete shells, see Becker (2013).

In summary, the expressions for the deflection angle and magnification matrix (equations 2 and 3) can be expressed as a sum over a discrete number of lensing planes such that at the n th plane they are given by

$$\beta^n = \beta^0 - \sum_{m=0}^{n-1} \frac{\chi^n - \chi^m}{\chi^n} \alpha^m, \quad (5)$$

and

$$A_{ij}^n = \delta_{ij} - \sum_{m=0}^{n-1} \frac{\chi^n - \chi^m}{\chi^n} U_{ik}^m A_{kj}^m, \quad (6)$$

where χ^m is the comoving distance to the m th plane. For clarity, we left out the dependence of the quantities on χ and θ . Here, we defined the shear matrix U such that the deflection field (α^m) and shear matrix equal the first- [$\alpha^n(\theta) = \nabla_{\perp} \psi^n(\theta)$] and second-order ($U_{ij}^n = \partial_i \partial_j \psi^n$) derivatives, orthogonal to the direction of propagation, of the lensing potential ψ , which is defined as

$$\psi^n \equiv \frac{2}{\chi^n c^2} \int_{\chi_{\min}}^{\chi_{\max}} d\chi \Phi, \quad (7)$$

where χ_{\min} and χ_{\max} are the comoving distances to the beginning and end of the shell, respectively. In principle, using equations (5) and (6), the magnification matrix and angular positions at the next plane are evaluated using those of all previous planes. It is, however, computationally infeasible to construct full-sky maps with a resolution sufficiently high for the analysis of Stage IV WL surveys in this way. Hilbert et al. (2009) showed that the sum can be combined into a recurrence relation where only the information of the previous two shells needs to be stored in memory. The recurrence relation is a result of the exact form of the transverse comoving distance in a

generic Robertson–Walker metric and is given by (Schneider 2016):

$$A_{ij}^{n+1} = \left(1 - \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}}\right) A_{ij}^{n-1} + \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}} A_{ij}^n - \frac{\chi^{n+1} - \chi^n}{\chi^{n+1}} U_{ik}^n A_{kj}^n, \quad (8)$$

with the initialization given by $A_{ij}^{-1} = A_{ij}^0 = \delta_{ij}$. A similar recurrence relation holds for the angular position, where the positions at the next plane (β^{n+1}) can be evaluated using the position at the current plane (β^n) and the previous plane (β^{n-1}),

$$\beta^{n+1} = \left(1 - \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}}\right) \beta^{n-1} + \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}} \beta^n - \frac{\chi^{n+1} - \chi^n}{\chi^{n+1}} \alpha^n, \quad (9)$$

where $\beta^{-1} = \beta^0 = \theta$ are the initial angular positions of the rays. In Section 3.4, we will explain how the FLAMINGO light-cones can be related to the deflection field and shear matrix in equations (8) and (9). The equations can then be combined with equation (4) to estimate the WL convergence, κ .

3. METHODS

3.1 The FLAMINGO simulations

For our analysis, we make use of the FLAMINGO simulation suite, a recent collection of large cosmological hydrodynamical simulations explicitly designed for the purpose of LSS analysis and cluster physics. For a full description of the simulation details, its performance with respect to observables, and calibration strategy see Schaye et al. (2023), Kugel et al. (2023), and McCarthy et al. (2023). We summarize here the key elements.

The simulations were run using the SWIFT hydrodynamics code (Schaller et al. 2023) with the SPHENIX smoothed particle hydrodynamics implementation (Borrow et al. 2022). Neutrinos are modelled as massive particles, using the δf method of Elbers et al. (2021) that was designed to reduce particle shot noise. The simulations include radiative cooling and heating that is implemented on an element-by-element basis (Ploekinger & Schaye 2020), star formation (Schaye & Dalla Vecchia 2008), and time-dependent stellar mass-loss as described by Wiersma et al. (2009). SN and stellar feedback are implemented kinetically (Dalla Vecchia & Schaye 2008) by kicking neighbouring particles in a way that conserves energy as well as linear and angular momentum, as described by Chaikin et al. (2023). The accretion of gas on to supermassive black holes (BHs) and the subsequent thermal AGN feedback is described in Booth & Schaye (2009) and the kinetic jet feedback is based on the AGN jet implementation of Huško et al. (2022), where gas particles receive a kick to a fixed target jet velocity in the direction given by the spin of the BH.

An important novel feature in the suite is that the runs have been calibrated using Gaussian process emulators trained on Latin hypercubes of smaller simulations where subgrid parameters are varied (Kugel et al. 2023). In this way, to study the impact of feedback variations, instead of varying a single, unobservable parameter relating to the specific implementation of a feedback process, a set of subgrid parameters is systematically varied by fitting a Gaussian process emulator such that the simulations can be characterized by (shifts in) real observables instead of subgrid parameters. The FLAMINGO variations were calibrated to the observed $z = 0$ galaxy SMF and gas fractions in low- z clusters. Comparing sets

Table 1. Characteristics of the four simulation variations that differ only in numerical resolution and box size. The columns indicate the simulation identifier; the box size length in cGpc, L ; the number of baryonic and DM particles, N_b ; the number of massive neutrino particles, N_ν ; the initial mean baryonic particle mass, m_b ; and the mean CDM particle mass, m_{CDM} .

Identifier	L cGpc	N_b	N_ν	m_b M_\odot	m_{CDM} M_\odot
L1_m8	1	3600 ³	2000 ³	1.38×10^8	7.06×10^8
L1_m9	1	1800 ³	1000 ³	1.07×10^9	5.65×10^9
L1_m10	1	900 ³	500 ³	8.56×10^9	4.52×10^{10}
L2p8_m9	2.8	5040 ³	2800 ³	1.07×10^8	5.65×10^8

of simulations calibrated with different observables allows for more instructive comparisons than comparing simulations that differ in specific subgrid parameters. Additionally, expected observational biases were included in the calibration. For the study of WL peaks, which trace the total mass in objects, calibrating to these observables ensures that the objects causing the peaks have a realistic ratio of gas and stars to DM. In the suite, the four subgrid parameters that are varied relate to the subgrid prescription of the SN and AGN feedback, namely the fraction of stellar energy feedback that couples to the interstellar medium (f_{SN}), the target wind velocity for SN feedback (Δv_{SN}), the BH accretion rate boost factor (β_{BH}), and either the AGN heating temperature (ΔT_{AGN}) or jet velocity (v_{jet}) for thermal and kinetic jet AGN feedback, respectively.

At fixed fiducial feedback calibration, the suite has four different runs varying the mass resolution and box size. The details of these runs are listed in Table 1. The identifier of each run indicates the box size in comoving Gpc (cGpc) and the rounded \log_{10} mass of the baryonic particle mass. For example, the flagship run (L2p8_m9) is a 2.8 cGpc box with 0.3 trillion (5040³ for DM and baryons and 2800³ for massive neutrinos) particles with a baryonic particle mass of $1.07 \times 10^9 M_\odot$, making it the cosmological hydrodynamical simulation with the highest number of resolution elements run to $z = 0$ to date. For each of the models, a corresponding DMO+ ν run, with the same initial phases, exists, whose identifier carries the post-fix ‘_DMO’. The cosmology of these runs was taken from the Dark Energy Survey year 3 ‘3 \times 2pt + All Ext.’ Λ CDM cosmology (Abbott et al. 2022), indicated as ‘D3A’ in Table 3. The initial conditions were generated using MONOFONIC (Hahn et al. 2020; Michaux et al. 2021) using three-fluid third-order Lagrangian perturbation theory with separate transfer functions for baryons, cold dark matter (CDM), and neutrinos (Elbers et al. 2022). The simulations are initiated at a redshift of $z = 31$.

3.2 Model variations

The suite includes 12 model variations in L1 boxes. The details of all the runs varying the baryonic feedback model with respect to the fiducial 1 cGpc box are listed in Table 2. At fixed cosmology, eight runs were calibrated to different galaxy SMFs (M_*) and/or gas fractions in clusters (f_{gas}) and/or differ in overall AGN subgrid feedback prescription. The subgrid parameter values have been chosen such that the SMF and/or gas fraction in clusters within the simulation are a set number of standard deviations from the fiducial model (ΔM_* and Δf_{gas} , respectively), as indicated in the second and third columns of Table 2. The standard deviation on the gas fractions was estimated by bootstrapping the X-ray data (table 5 of Kugel et al. 2023) and the error on the WL data (Akino et al. 2022). Similarly, the shift of the SMF is the expected systematic error (0.14 dex) on the stellar masses from Behroozi et al. (2019).

Table 2. The baryonic physics variations in the 1 cGpc FLAMINGO box. For each of the different models, indicated by their identifier, the second and third columns indicate the number of observational standard deviations (σ) by which the galaxy SMF (M_*) and gas fractions in clusters (f_{gas}) in the simulation are shifted compared with the fiducial L1_m9 model, respectively. The final column indicates the method of AGN feedback in the run. All the runs with different baryonic feedback were run using the same D3A cosmology (see Table 3) and with the same initial conditions.

Identifier	ΔM_* σ	Δf_{gas} σ	AGN mode
L1_m9	0	0	Thermal
$f_{\text{gas}}+2\sigma$	0	+2	Thermal
$f_{\text{gas}}-2\sigma$	0	-2	Thermal
$f_{\text{gas}}-4\sigma$	0	-4	Thermal
$f_{\text{gas}}-8\sigma$	0	-8	Thermal
$M_*-\sigma$	-1	0	Thermal
$M_*-\sigma, f_{\text{gas}}-4\sigma$	-1	-4	Thermal
Jet	0	0	Jet
Jet $f_{\text{gas}}-4\sigma$	0	-4	Jet

The suite includes more models with stronger than weaker feedback to allow quantifying the observational signatures of exceptionally strong feedback, which affects larger length scales than exceptionally weak baryonic feedback.

There are four runs varying only the gas fractions in clusters, whose identifiers are given by ' $f_{\text{gas}} \pm n\sigma$ ', where n is the number of standard deviations by which the gas fraction is varied. One run was calibrated to match the galaxy SMF shifted to a lower mass by 1σ ($M_*-\sigma$). The $M_*-\sigma, f_{\text{gas}}-4\sigma$ run varies both observables. The fiducial implementation of AGN feedback is thermal, but there are two models with a kinetic jet AGN feedback description, which are denoted by 'Jet' and 'Jet $f_{\text{gas}}-4\sigma$ ', where for the latter the target gas fraction in clusters is reduced. The jet models can be used to access the sensitivity of observables to variations in subgrid models calibrated to the same observables.

At fixed baryonic calibration, there are four cosmology variations. The different cosmologies and their cosmological parameters are listed in Table 3. In addition to the fiducial D3A cosmology that was used for all the different baryonic physics runs, there are three variations based on CMB measurement from Planck. The first CMB cosmology is the best-fitting Planck Collaboration VI (2020) Λ CDM cosmology with $\sum m_\nu = 0.06$ eV (Planck). The other two Planck cosmology variations include heavier neutrinos (three species each with $m_\nu c^2 = 0.08$ eV), one in which the other parameters are changed according to their best-fitting values within the Planck Markov Chain Monte Carlo chains (PlanckNu0p24Var), and one in which the other parameter values are fixed and only Ω_{CDM} is adjusted to keep Ω_m

fixed (PlanckNu0p24Fix). The final variation is a cosmology model with a lower value of S_8 (LS8), taken from Amon et al. (2023).

Critically, all variations in the L1 box have been run using the same initial phases. This allows us to isolate the effect of the varying baryonic physics and cosmology from cosmic variance induced by different initial realizations.

3.3 Light-cones

To construct the WL convergence maps, we use the mass maps of the FLAMINGO light-cones. The light-cones correspond to virtual observers within the FLAMINGO simulation suite. They have been constructed by recording particles crossing the past light-cone of a virtual observer. The light-cones of the L1 runs are provided as 60 projected shells spaced equally in redshift between $z = 0$ and 3 (i.e. $\Delta z = 0.05$). Particles within the innermost 3 Mpc have been removed. The shells are stored as HEALPix maps (Górski et al. 2005) with a resolution of $N_{\text{side}} = 16384$ (with a number of pixels of $N_{\text{pix}} = 12N_{\text{side}}^2$), corresponding to an angular resolution of 0.21 arcmin. For the details of the light-cone construction and properties, see appendix A of Schaye et al. (2023).

As the box size is not large enough to cover the distance up to $z = 3$, which is the range we consider (see Section 3.4.2), the box is replicated to reach this distance. For the L2p8 and L1 boxes, this requires 5 and 12 additional replications, respectively. In Appendix A, we compare different ways of implementing the box replication, including different mass shell rotation strategies. We compare the signal in the L1_m10.DMO run using non-rotated shells, the case in which every shell is randomly rotated and correlations along the line of sight may thus be unnecessarily erased, and the case of rotating the shells whenever the light-cone diameter is larger than the box length. We compare the measurements to those for a 5.6 cGpc DMO run (L5p6_m10.DMO), which can cover the light-cone diameter without replications up to $z = 0.8$. We find only minor differences between the different box rotation strategies. To prevent encountering the same structure multiple times, but also to not unnecessarily erase correlations along the line of sight, we choose to randomly rotate the shells every half box length. The same random angles are applied to all L1 observers. As these observers are all placed at the same position and reside in boxes with the same initial phases, we can directly study the impact of both the cosmology variations and baryonic physics on the measured WL signal generated by the same objects. Within the L2p8_m9 box, eight observers have been placed at the coordinates $(\pm L/4, \pm L/4, \pm L/4)$, where L is the simulation box size. These light-cones have 68 shells up to $z = 5$ but we only use the first 60 light-cones to facilitate a direct comparison with the L1 light-cones, as the first 60 shells have

Table 3. The cosmological parameter values used in the different FLAMINGO simulations. The columns indicate the cosmology identifier; the dimensionless Hubble constant, $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$; the total matter density parameter, Ω_m ; the baryonic matter density parameter, Ω_b ; the cosmological constant density parameter, Ω_Λ ; the neutrino matter density parameter, Ω_ν ; the summed mass of the massive neutrino species, $\sum m_\nu c^2$; the amplitude of the primordial power spectrum, A_s ; the power-law index of the primordial matter power spectrum, n_s ; the amplitude of the linear theory power spectrum parametrized as the rms mass density fluctuation in spheres of radius $8 h^{-1} \text{ Mpc}$ at $z = 0$, σ_8 ; and the amplitude of the initial power spectrum parametrized as $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3}$.

Cosmology	h	Ω_m	Ω_b	Ω_Λ	Ω_ν $\times 10^{-3}$	$\sum m_\nu c^2$ (eV)	A_s $\times 10^{-9}$	n_s	σ_8	S_8
D3A	0.681	0.306	0.0486	0.694	1.39	0.06	2.099	0.967	0.807	0.815
Planck	0.673	0.316	0.0494	0.684	1.42	0.06	2.101	0.966	0.812	0.833
PlanckNu0p24Var	0.662	0.328	0.0510	0.672	5.87	0.24	2.109	0.968	0.772	0.807
PlanckNu0p24Fix	0.673	0.316	0.0494	0.684	5.69	0.24	2.101	0.966	0.769	0.789
LS8	0.682	0.305	0.0473	0.695	1.39	0.06	1.836	0.965	0.760	0.766

the same redshift spacing as the light-cones of the L1 runs. The mean shell thickness is ≈ 110 Mpc. Zorrilla Matilla, Waterval & Haiman (2020) found that the bias resulting from such a redshift discreteness is statistically insignificant in the presence of shape noise for Stage IV cosmic shear statistics. We compare the eight different light-cones to quantify the impact of cosmic variance on our analysis. Additionally, we carry out numerical convergence tests for simulation box size and resolution.

3.4 Full-sky WL convergence maps

3.4.1 Spherical harmonics

To exploit the full power of the all-sky FLAMINGO HEALPix maps, we carry out our analysis in spherical harmonics space. Because some Fortran compilers do not support 64-bit array sizes, the HEALPix library officially does not support maps with a size larger than $N_{\text{side}} = 8192$. Even though our analysis uses the Python implementation of the HEALPix library (Zonca et al. 2019), which is technically not affected by this issue, some internal functionality that we need relies on 32-bit indexing, which means we cannot use these functions on the full-resolution FLAMINGO light-cone maps. We therefore limit our analysis to downsampled maps with $N_{\text{side}} = 8192$, corresponding to a pixel size of 0.43 arcmin for a full-sky analysis. As we smooth our final maps with a Gaussian kernel with a full width at half-maximum (FWHM) of 1 arcmin, we are not directly limited by the resolution of the HEALPix maps, which we illustrate in more detail in Section 3.5. At each of the light-cone shells, we determine the lensing potential (given by equation 7) using the solution of the two-dimensional (2D) Poisson equation which we define as the convergence at plane n (K^n) as

$$K^n(\theta) \equiv \nabla^2 \psi^n = \frac{3H_0^2 \Omega_m}{2c^2} \chi^n (1 + z^n) \Delta \chi^n \delta^n(\theta), \quad (10)$$

where $\Delta \chi = \chi_{\text{max}} - \chi_{\text{min}}$ is the thickness of the shell and we choose to evaluate χ^n and z^n at the comoving centre of each shell. The overdensity $\delta^n(\theta)$ can be directly evaluated from the surface mass density and is given by

$$\delta^n(\theta) = \frac{\Sigma^n(\theta) - \bar{\Sigma}^n}{\bar{\Sigma}^n}, \quad (11)$$

where $\Sigma^n(\theta)$ is the surface density at position θ and $\bar{\Sigma}^n$ is the mean surface density of the n th shell for the given cosmology, which we evaluate directly from the shell.

To exploit the full-sky light-cones, the equation for the lensing potential is solved in spherical harmonics space (e.g. Hu 2000; Price et al. 2021). The convergence at plane n is related to the lensing potential using the spherical harmonics coefficients of the convergence ($K_{\ell m}^n$) and lensing potential ($\psi_{\ell m}^n$) as

$$\psi_{\ell m}^n = \frac{-2K_{\ell m}^n}{\ell(\ell + 1)}, \quad (12)$$

which can then be used to determine the derivatives necessary to compute the deflection field and shear matrix at each plane. As the mass maps are provided as discrete maps on a HEALPix grid, we need to adopt a strategy for determining the shear matrix and deflection field. For the lowest redshift shell, the rays can be conveniently ‘aimed’ directly at the pixel centres. Since the photons are deflected here, their paths will, in general, not pass through the centre of a pixel in a subsequent shell. Therefore, the magnification matrix and the deflection angle for each ray are evaluated using bilinear interpolation as the weighted average of the four nearest pixels,

similar to Shirasaki, Hamana & Yoshida (2015). We have found that the bilinear interpolation introduces some smoothing of the signal, but overall the additional smoothing of the final maps (see below) dominates. Throughout the recurrence relations (equations 8 and 9), we determine the lensing potential using equation (12). The lensing potential is then converted to the deflection field (α) and shear matrix (U_{ij}) through its first- and second-order covariant derivatives, which are used to calculate the magnification matrix (A) and the angular position (β) at each plane, where we take into account the change in basis as the photon gets displaced (Becker 2013).

3.4.2 Source redshift distribution

Following the above procedure, the magnification matrix is determined for all rays at all planes. The magnification matrix can then be related to the WL convergence (κ) via equation (4). To mimic a signal corresponding to a Stage IV survey, we incorporate a source redshift distribution [$n(z)$] as

$$\kappa(\theta) = \int_0^{\chi_{\text{hor}}} d\chi n(z(\chi)) \kappa(\theta, \chi), \quad (13)$$

where the integral runs over the line of sight to the edge of the survey, which in our case reduces to a discrete sum over the shells until $\chi_{\text{hor}} = \chi(z = 3)$. Here, we use a simple *Euclid* mock forecast given by (Euclid Collaboration 2020):

$$n(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right], \quad (14)$$

with $z_0 = 0.9/\sqrt{2}$. The function is normalized such that $\int n(z) dz = 1$, and relates to the comoving distance as $n(z) dz = n(\chi) d\chi$. More realistic alternatives to this distribution exist, but for our purposes, this simple distribution suffices as we do not expect minor consistent changes in the redshift distribution to impact the overall baryonic effects. The source redshift distribution is shown in Fig. 1. The top axis indicates the comoving distance corresponding to our fiducial cosmology. Additionally, the green, blue, and black arrows indicate the box sizes of the L1, L2p8, and L5p6 variations, respectively.

3.4.3 Smoothing and noise

To mimic an observed signal, we add galaxy shape noise to each pixel of the final map by drawing from a normal distribution with mean μ and standard deviation σ (Kaiser & Squires 1993; Lu & Haiman 2021):

$$\mathcal{N}\left\{\mu = 0, \sigma = \frac{\sigma_\epsilon}{\sqrt{2n_{\text{gal}}A_{\text{pix}}}}\right\}, \quad (15)$$

where σ_ϵ is the rms intrinsic ellipticity of source galaxies, n_{gal} is the source number density on the sky, and A_{pix} is the pixel area. In our case, we choose $\sigma_\epsilon = 0.26$ and $n_{\text{gal}} = 30 \text{ arcmin}^{-2}$, to mimic the expected signal that *Euclid* will measure (Laureijs et al. 2011). This σ_ϵ value has been measured from *Hubble Space Telescope* images with similar photometric properties as the expected *Euclid* images and is therefore commonly used to model the observed galaxy shape noise (Schraback et al. 2018; Euclid Collaboration 2019, 2023). The final maps are smoothed by a Gaussian kernel with a FWHM of 1 arcmin. This was identified by Liu et al. (2015) as an optimal smoothing scale to counterbalance the loss of cosmological information and the minimization of noise. In our case, this corresponds to ≈ 2.5 pixels.

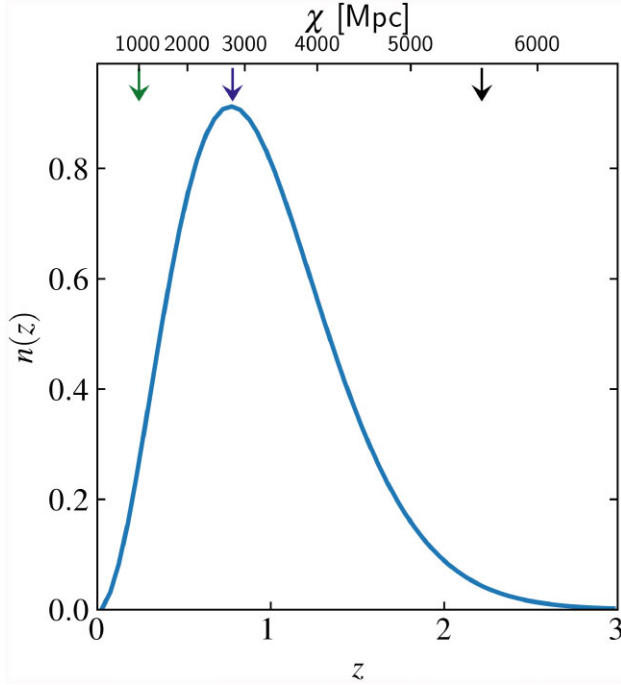


Figure 1. *Euclid*-like source redshift distribution used in the construction of the WL convergence maps. The top axis shows the comoving distance for the fiducial D3A cosmology. The green, blue, and black arrows indicate the box size of the L1 (1 cGpc), L2p8 (2.8 cGpc), and L5p6 (5.6 cGpc) variations, respectively.

Fig. 2 shows a full-sky WL convergence map corresponding to an observer in the L2p8.m9 simulation run. Here, the full-sky map and top $5 \times 5 \text{ deg}^2$ zoom-in have not been smoothed and no noise has been applied. The bottom zoom-in shows the same area but corresponds to the final WL convergence map where the noise and smoothing have been applied. The highest valued peaks are visible in both panels but the noise creates spurious low signal-to-noise peaks.

3.4.4 Second-order effects

Our analysis is somewhat idealized as we ignore several second-order effects such as lens–lens coupling (Bernardeau, van Waerbeke & Mellier 1997), and the fact that the true cosmic shear observable is the reduced shear $g = \gamma/(1 - \kappa)$. As explained in Kilbinger (2015), most of these second-order effects are at least two orders of magnitude smaller than the first-order convergence angular power spectrum. The dominant contribution originates from the reduced shear correction, which contributes up to 10 per cent of the signal on arcmin scales. However, in our analysis, we assume that the reconstruction of the signal takes into account the reduced shear correction, which can, e.g. be done using a quickly converging iteration (Bradač et al. 2005).

3.5 Angular power spectra

To validate the construction of our WL convergence maps and assess their robustness, we compute the angular power spectrum, $\mathcal{C}(\ell)$, for the L2p8.m9.DMO run for different choices of HEALPix grid resolution, smoothing, and ray-trace methodology. The angular power spectra are computed from the WL convergence map using the HEALPix ANAFast routine and are shown in Fig. 3. We compare with the prediction from HALOFIT (Takahashi et al. 2020) using

the CLASS code (Blas, Lesgourgues & Tram 2011). HALOFIT is a commonly used fitting formula to the (non-linear) 3D matter power spectrum (P_m) based on high-resolution N -body simulations. We relate the matter power spectrum to the angular power spectrum using (Limber 1953; LoVerde & Afshordi 2008):

$$\mathcal{C}(\ell) = \int_0^{\chi_{\text{hor}}} d\chi \frac{W^2(\chi)}{\chi^2} P_m\left(\frac{\ell + 1/2}{\chi}, z(\chi)\right), \quad (16)$$

where we assume the fiducial D3A cosmology. $W(\chi)$ is the WL kernel which is given by (Kaiser 1992):

$$W(\chi) = \frac{3H_0^2\Omega_m}{2c^2} \chi (1 + z(\chi)) \int_{\chi}^{\chi_{\text{hor}}} d\chi' n(\chi') \frac{\chi' - \chi}{\chi'}. \quad (17)$$

The HALOFIT prediction is given by the dash–dotted green curve in Fig. 3. The dashed red curve corresponds to a ray-traced convergence map where the quantities that are determined for each photon at every shell (i.e. the deflection angle and shear matrix), were determined using the value of the nearest grid point (NGP) for each ray. In this case, there is excellent agreement between the ray-traced angular power spectrum and the theoretical prediction up to $\ell \approx 6000$. The deviation at smaller scales (larger ℓ) is a result of the pixelation of the HEALPix grid. Also, as illustrated by Upadhye et al. (2023), the HALOFIT prediction deviates from FLAMINGO by a few per cent at small scales.

The dotted red curve corresponds to the same map as the dashed red curve but downsampled to a lower resolution of $N_{\text{side}} = 4096$, corresponding to an angular resolution of 0.86 arcmin. As expected, the deviation from the high resolution and HALOFIT predictions begins at smaller ℓ and it is greater in magnitude when the HEALPix grid has a coarser resolution. The dashed black curve corresponds to the case where the quantities at each shell are determined using bilinear interpolation. Because the quantities are determined as a weighted average of the four nearest neighbours on the 2D grid, some de facto smoothing at the pixel scale is applied at each shell. We find that this introduces some loss of power on small scales, beyond $\ell \approx 2 \times 10^3$, to a similar degree as downsampling the resolution of the HEALPix map to $N_{\text{side}} = 4096$. Both the loss of power on small scales due to the smoothing introduced by the interpolation, and that due to the pixelation of the HEALPix grid, are several times smaller than the suppression introduced by smoothing the final maps with a Gaussian kernel with an FWHM of 1 arcmin, as illustrated by the solid red and black curves for NGP and bilinear interpolation, respectively.

Over all scales, smoothing dominates all the choices that were made in the construction of the maps. However, there is still some additional power loss when using bilinear interpolation compared with NGP interpolation. At $\ell = 2000$ ($\theta \approx 5$ arcmin), the difference between the interpolation strategies is 1 per cent and it increases for smaller scales. We nevertheless choose to apply bilinear interpolation, which is also used in the ray-tracing method of Hilbert et al. (2009) and which was shown by Hilbert et al. (2020) to have a peak distribution for peaks with a signal-to-noise ratio (SNR) < 6 that agrees to within 5 per cent with codes that either use higher resolution equidistant cylindrical projection pixelization or determine the convergence signal on the fly (Muciaccia, Natoli & Vittorio 1997; Barreira et al. 2017; Fabbian, Calabrese & Carbone 2018). Ferlito et al. (2023) have studied the convergence with the pixel size at the same smoothing scale that we use and they find that the distributions are suppressed by up to a couple per cent compared with maps with a higher resolution. We do not expect these few per cent differences to impact our main conclusions as we consistently apply the same approach to all simulation variations. Thus, we still get robust estimates of the differences in number densities of peaks. However,

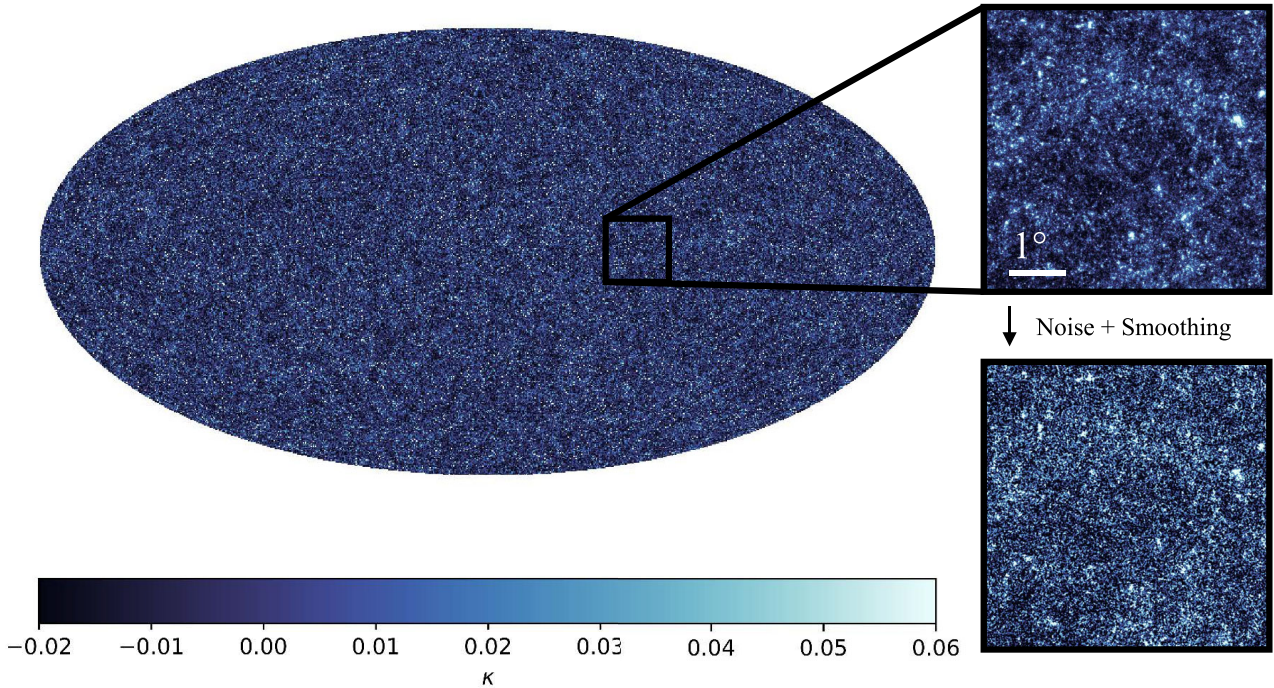


Figure 2. Ray-traced full-sky WL convergence (κ) map (left panel) and zoom-ins of a $5 \times 5 \text{ deg}^2$ region for a *Euclid*-like source redshift distribution as measured by a virtual observer in the hydrodynamical L2p8_m9 run. The full-sky map and top zoom-in correspond to a WL convergence map without smoothing or noise. The bottom zoom-in contains galaxy shape noise and is smoothed with a Gaussian kernel with FWHM = 1 arcmin, both of which are applied before determining the WL peaks. The map is generated using a backward ray-tracing method where the photons that are observed by the virtual observer are deflected by the matter distribution at 60 linearly spaced discrete lensing planes between $z = 0$ and 3.

to eventually compare to observations, it might be necessary to use HEALPix maps with a higher angular resolution, whilst sticking to a 1-arcmin smoothing, to avoid sensitivity to the interpolation scheme and to avoid sensitivity to the discretization of the HEALPix grid.

4. RESULTS

In this section, we start by quantifying the level of numerical convergence in terms of resolution and box size, and the effect of cosmic variance. We then compare the signals from the different baryonic physics models and the runs with different cosmological models. Finally, we study the separability of the cosmological and baryonic impact. We quantitatively and qualitatively compare our results with previous studies. In our analysis, we quantify the distribution of peaks in the WL convergence maps as a function of their κ value. A peak is defined as any pixel that has a value larger than those of the eight closest pixels on the HEALPix grid. We use the number density of peaks to study the peak distributions. We define the number density of peaks ($dn/d\kappa$) as the number of peaks per square degree divided by the convergence bin size.¹

¹Although the quantity is independent of bin size, for reference, we state the bin edges rounded to three decimal places [−0.100, −0.085, −0.069, −0.054, −0.040, −0.025, −0.020, −0.015, −0.010, −0.004, 0.001, 0.006, 0.011, 0.016, 0.021, 0.026, 0.032, 0.037, 0.042, 0.047, 0.052, 0.057, 0.062, 0.068, 0.073, 0.078, 0.083, 0.101, 0.153, 0.204, 0.256, 0.308, 0.323, 0.500, 0.563, 1.000]

4.1 Numerical convergence

As discussed in Section 3.2, all cosmological and baryonic feedback variations are carried out in L1 boxes at the m9 resolution. Therefore, we first quantify the numerical convergence of the fiducial L1_m9 run with respect to box size (L) and resolution (m). To isolate convergence effects from baryonic effects, we study the numerical convergence in the accompanying DMO+ ν runs, where the convergence of results is well established and understood. To this end, the top panel in Fig. 4 shows the number density of WL peaks as measured for the observers in L1_m8_DMO (red), L1_m9_DMO (green), L1_m10_DMO (yellow), and the mean of the eight observers in L2p8_m9_DMO (blue). The high-resolution run has a $64 \times$ higher mass resolution than the low-resolution run. The top axis of the figure shows the $\text{SNR} = \kappa/\sigma$, where σ is the standard deviation of the smoothed galaxy shape noise map (equation 15). The bottom panel shows the ratio of the distribution of each model relative to that of L1_m9_DMO. The Poisson error for L1_m9_DMO is indicated by the shaded area in the bottom panel. (The noise does not increase monotonically for larger κ as the bin size is not kept constant in order to decrease the noise in large- κ bins.)

The differences between the different resolution L1 runs, which have the same initial phases and are thus not impacted by cosmic variance, can be used to assess the numerical convergence with resolution. The agreement between the fiducial run and the higher resolution run (L1_m8_DMO), as well as with the larger box size run (L2p8_m9_DMO), does not exceed 0.5(2) per cent up to $\kappa = 0.1$ (0.2) and is excellent up to $\kappa \approx 0.4$, illustrating that the measurement is well converged in this regime. These findings are consistent with

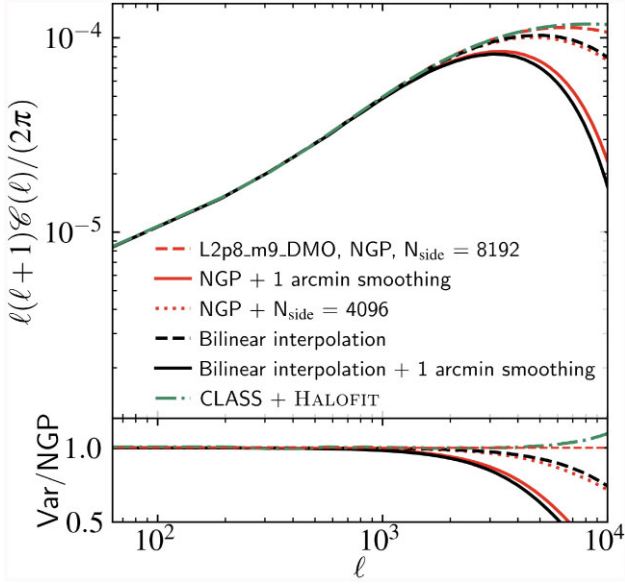


Figure 3. Top panel: WL convergence angular power spectrum for variations of smoothing, angular resolution, and interpolation strategy for a single observer in the L2p8.m9.DMO run and the theoretical non-linear prediction from CLASS + HALOFIT. Bottom panel: Ratios of the different curves in the top panel and the one for NGP interpolation at $N_{\text{side}} = 8192$ (dashed red curve). The 1-arcmin smoothing dominates over any suppression induced by pixelation or ray tracing but differences of a few percent are still present between different procedures. We use bilinear interpolation at $N_{\text{side}} = 8192$ with 1-arcmin smoothing (solid black curve) as our fiducial map construction method.

Ferlito et al. (2023) who studied the numerical and angular resolution convergence of the number density of WL peaks. For larger κ values, the deviations between the runs become larger than 10 percent. However, the number of peaks in this regime also drops and the distribution thus becomes dominated by Poisson noise, as shown by the grey-shaded region. Therefore, we do not expect the peaks in this regime to be cosmologically informative in practice. The lower mass resolution (L1.m10.DMO) shows a lower number of peaks for $\kappa > 0.1$, indicating the number density of WL peaks is not yet converged for resolutions lower than the fiducial m9. As the initial phases of the L1 runs are the same, the same haloes exist in these simulations. At $\kappa \gtrsim 0.4$, the variations between the number densities of the intermediate- and high-resolution runs are likely due to changes in the masses of individual haloes as the positions of the haloes should be unaltered, leaving the WL kernel unchanged. The L2p8.m9.DMO run has different initial conditions and is thus a different realization of the same Universe. The comparison with this run therefore suffers from cosmic variance, which we quantify in the next section. The L1.m8.DMO and L2p8.m9.DMO simulations show a slightly better degree of convergence than the L1.m9.DMO run. Especially up to $\kappa = 0.2$, the larger box and higher mass resolution runs agree almost perfectly. However, this agreement is fortuitous, because it is the result of box size and resolution compensating each other. The comparisons between the resolutions or box sizes illustrate that the signal for L1 and m9, which are used for all cosmological and baryonic variations, is well converged in the regime $\kappa \lesssim 0.4$. Unlike larger κ values, this regime is not impacted significantly by Poisson errors and cosmic variance (as shown in the next section), and will thus be most informative when inferring cosmological constraints.

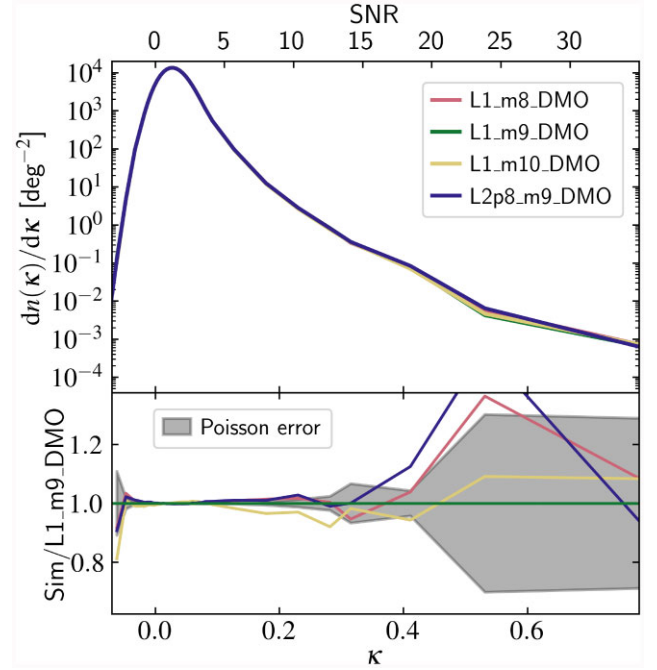


Figure 4. Top panel: Number density of WL peaks for the observers in the L1.m8, L1.m9, L1.m10, and L2p8.m9 DMO runs, which differ in terms of resolution and/or box size. The L2p8.m9.DMO signal is the mean of the eight observers in the box. The $\text{SNR} = \kappa/\sigma$ (top x-axis) is computed from the standard deviation of a smoothed noise realization. Bottom panel: Ratio to the L1.m9.DMO run. The shaded region indicates the Poisson error for L1.m9.DMO. The numerical convergence of the fiducial (m9) resolution and box size (L1), which was used for all the baryonic and cosmology variations, is excellent up to $\kappa = 0.3$ and adequate for larger values, where the Poisson error gets larger.

4.2 Cosmic variance

Using the eight independent observers within the L2p8.m9.DMO run, we can directly test the impact of cosmic variance on the measured statistics. The top panel of Fig. 5 shows the number density of peaks for each of the eight light-cones. The bottom panel shows their ratios with their mean. The standard deviation of the ratio is indicated by the grey-shaded region. For positive WL convergence values up to $\kappa = 0.1(0.2)$, the distributions agree to within 0.2(1) percent precision. Up to $\kappa \approx 0.4$, the difference is not greater than 10 percent, after which the degree of variation increases sharply. For $\kappa > 0.4$, the number density is so low that there exist fewer than 10^2 peaks within a convergence bin on the entire sphere. As the WL convergence signal is most sensitive to overdensities roughly halfway in between the observer and the source galaxy, the exact configuration of the observers with respect to the most massive haloes in the simulation will determine the number of peaks in the highest- κ bins (Kilbinger 2015). This is reflected in the distributions as the variance increases for larger WL convergence values.

Based on the comparisons in this and the previous section, we find that our measurements are robust up to at least $\kappa = 0.4$. For larger WL convergence values, the difference increases but both the Poisson noise as well as the uncertainty due to cosmic variance start to dominate the signal, which will be the limiting factor for any WL survey using the number density of high- κ WL peaks. The comparison in this and the previous section shows that the most informative regime for the number density of WL peaks is $\kappa \approx 0.1$ – 0.4 . The $\kappa > 0.4$ regime suffers from cosmic variance and Poisson

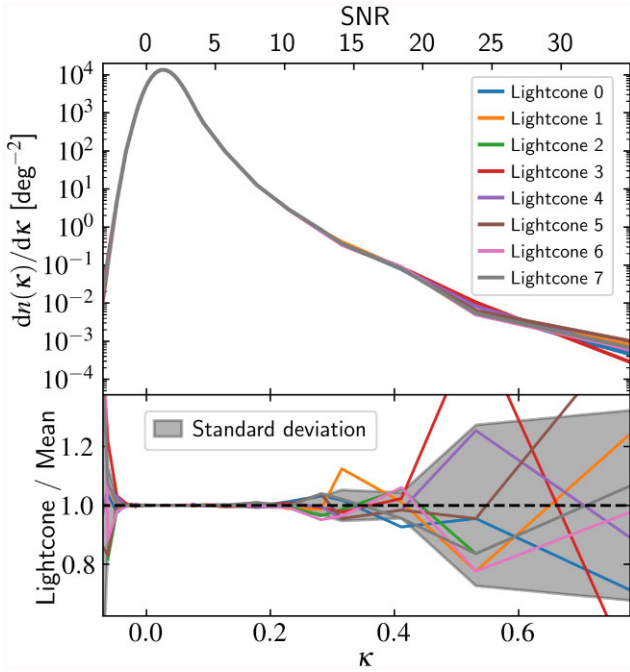


Figure 5. Top panel: Number density of WL peaks for eight different virtual observers in the L2p8_m9_DMO simulation. The observers were placed at the coordinates $(\pm L/4, \pm L/4, \pm L/4)$, where $L = 2.8$ cGpc is the simulation box size. Bottom panel: Ratio with the mean of the eight observers. The grey shading shows the standard deviation of the eight observers. Over the entire range of κ , the cosmic variance is comparable with the Poisson errors. At $\kappa > 0.4$, the uncertainty due to cosmic variance increases rapidly, which will be the limiting factor for any inference based on the number density of WL peaks.

noise and $\kappa < 0.1$ is impacted by the smoothing and noise, as is discussed more in the next section. Comparing the cosmic variance effect with the Poisson uncertainty in Fig. 4, we see that over the entire κ range, the two are of similar magnitude.

4.3 Comparison of fiducial DMO and hydrodynamical models

In this section, we explore the differences between the number density of WL peak distributions measured in a DMO and hydrodynamical run. To this end, we compare the mean of the measured signal of the eight virtual observers in the L2p8_m9 and L2p8_m9_DMO runs. Their mean number densities of peak distributions are shown in the top panel of Fig. 6 in blue and black, respectively. As the initial phases, observer positions, and applied rotations are the same, any difference is a direct measurement of the impact of baryonic physics on the WL convergence peak count distribution. The bottom panel shows the ratio of the mean of the hydrodynamical run compared with that of the DMO distributions. The dashed horizontal lines in the top panel indicate the regime at which we expect to measure only one peak in the *Euclid* footprint ($\approx 15\,000$ deg², Laureijs et al. 2011) and the KiDS footprint (≈ 1550 deg², Kuijken et al. 2019) with a larger WL convergence value. As *Euclid* will cover $\sim 1/3$ rd of the sky, close to the entire convergence regime can be probed. At small convergence values, $0 \lesssim \kappa \lesssim 0.05$, the baryonic physics enhances the number density of WL peaks by a few per cent. In Appendix B, we explore the impact of the applied smoothing and noise and illustrate that without the smoothing and noise, the hydrodynamic run shows a stronger enhancement of WL peaks around $\kappa = 0$. The comparison

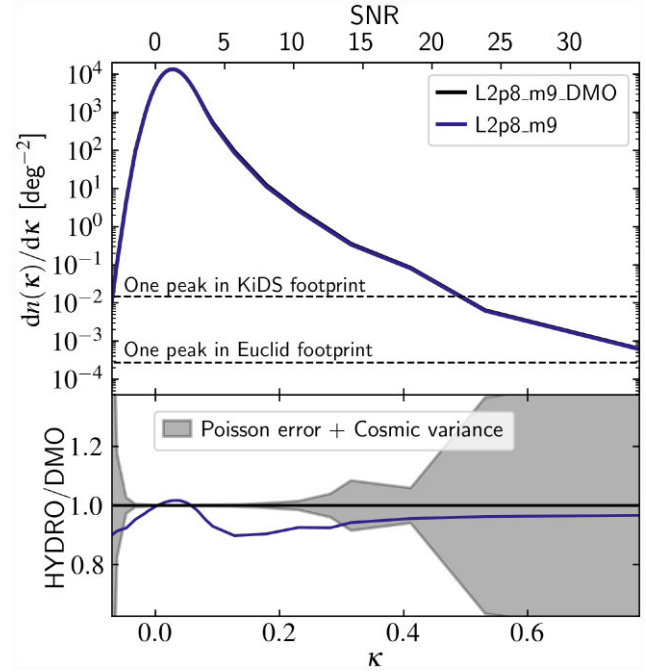


Figure 6. Top panel: Mean number density of WL peaks measured by the eight virtual observers in the DMO and cosmological hydrodynamical L2p8_m9 runs with the same cosmology and initial phases. The observers are located at the same positions in the simulations and we apply the same random rotations to their light-cone shells. The intersection of the dashed horizontal lines with the number densities indicates the κ regime where we expect to measure only a single peak with a larger κ value in the KiDS or *Euclid* footprints. Bottom panel: Ratio of the mean number density of peaks for the observers in the hydrodynamical run to the mean in the DMO run. The grey-shaded area indicates the quadrature sum of the estimate Poisson error and cosmic variance. The fiducial baryonic impact is largest at $\kappa \approx 0.1$, showing a 10 per cent suppression of the number density of WL peaks. The suppression due to baryons decreases almost monotonically for larger κ .

shows that the entire κ regime is impacted by the smoothing and noise and illustrates that for an actual inference, a more dedicated study that models these effects properly must be carried out.

At $\kappa \gtrsim 0.1$, there is a clear difference between the models, where we see a larger number of peaks in the DMO run. At $\kappa \gtrsim 0.4$, the number of peaks becomes very small and although the trend is still clear, this regime is dominated by Poisson noise and cosmic variance, as shown in Sections 4.1 and 4.2. The grey-shaded area corresponds to the quadrature sum of the Poisson error and cosmic variance as estimated for L2p8_m9_DMO, which will thus dominate over the baryonic impact for $\kappa > 0.4$.

We can understand the baryonic suppression of the peak distribution by considering differences for individual haloes between the hydrodynamical and DMO runs. Baryonic feedback will cause gas to be expelled from a halo causing the halo to be less massive (e.g. Velliscig et al. 2014; Bocquet et al. 2016), which directly decreases its lensing potential. Also, when gas is expelled early in a halo's lifetime, it will be less massive and consequently have a less deep gravitational potential and thus attract less matter over its entire evolution, increasing the mass difference compared with the DMO run (Stanek, Rudd & Evrard 2009; Cui et al. 2012). Debackere, Hoekstra & Schaye (2022) compared the halo masses of matched haloes in BAHAMAS and its DMO counterpart and showed that for haloes of mass $M_{200m} \approx [10^{13} - 10^{14}] M_{\odot}$, the halo mass in the hydrodynamical simulation is ~ 10 per cent lower than in the

DMO run. As the WL convergence signal is directly proportional to the overdensity (equation 10) and thus mass, we expect to observe weaker WL signals in the hydrodynamical run. The reduction in mass in hydrodynamical runs is directly reflected in the observed number densities, where we see a suppression of a similar magnitude. As shown by Debackere et al. (2022), for even more massive haloes ($M_{200m} \gtrsim 10^{14} M_\odot$), the feedback is not strong enough to effectively remove large amounts of gas from the haloes, and the mass difference between the haloes in different runs decreases. The same patterns are visible in the FLAMINGO HMFs reported in fig. 20 of Schaye et al. (2023), where compared with the DMO counterpart, the HMF is suppressed in the intermediate halo mass regime ($M_{200m} \approx [10^{13} - 10^{14}] M_\odot$) by 10–20 per cent, with larger differences for less massive haloes, but for the most massive haloes ($M_{200m} \approx 10^{15} M_\odot$) the suppression vanishes. We observe a similar effect as we see a decreasing suppression of the number density of WL peaks for $\kappa > 0.1$. Although the differences in halo mass and WL peak abundances agree quantitatively, we cannot conclude that haloes of specific masses are primarily responsible for the peaks. The WL convergence signal is also sensitive to the orientation of the observer with respect to the haloes and a peak with a larger κ value will thus not necessarily correspond to a more massive halo. However, in general, peaks with a larger κ value are more likely to originate from more massive haloes so qualitatively we do expect to see less suppression for larger κ values as we understand these peaks to originate from a single halo along the line of sight and the mass of the most massive haloes should be similar in the DMO and hydrodynamical runs (Yang et al. 2011). We aim to study the contribution of haloes of a specific mass range in future research.

We now compare our results with those from previous studies. We note that differences between our and previous analyses may arise because of different baryonic feedback implementations, source redshift distributions, and WL convergence map construction algorithms, which rely on choices for shape noise, smoothing, and ray tracing.

Coulton et al. (2020) reported on the impact of baryons on peak counts for a Rubin-like inference using the BAHAMAS simulations. Their analysis, which only extends to $\text{SNR} = 6$, also includes noise and smoothing, but their smoothing scale is twice as large as ours. In this convergence regime ($\kappa \lesssim 0.1$), which we found was most impacted by the applied noise and smoothing, they report a baryonic suppression that is roughly half of the suppression we measure. The reason for the difference is unclear, as the BAHAMAS runs were calibrated to match the same observables, but it illustrates that the magnitude of the baryonic suppression may depend on the details of the subgrid prescriptions in the simulation as well as the choices made in the construction of the convergence maps. Osato, Liu & Haiman (2021) have carried out a similar analysis, including a full ray-tracing treatment, in Illustris TNG (Springel et al. 2018) and they too found a suppression of the peak counts. Whereas they considered a single-redshift source sample and their smoothing scale is twice as large, they report a similar suppression factor. We cannot compare our results directly with the ones extracted by Ferlito et al. (2023) from the MTNG simulation, who studied convergence peaks in a suite of different hydrodynamical simulations, as their analysis only focuses on convergence values up to $\kappa = 0.06$ and they do not apply any observationally-inspired shape noise. In our case, this regime is completely dominated by the applied smoothing and shape noise.

In this section, we have shown that the fiducial baryonic feedback within the FLAMINGO simulation suite suppresses the number density of WL peaks as measured for a Stage IV WL survey by 10 per cent, and that the suppression decreases for larger κ values.

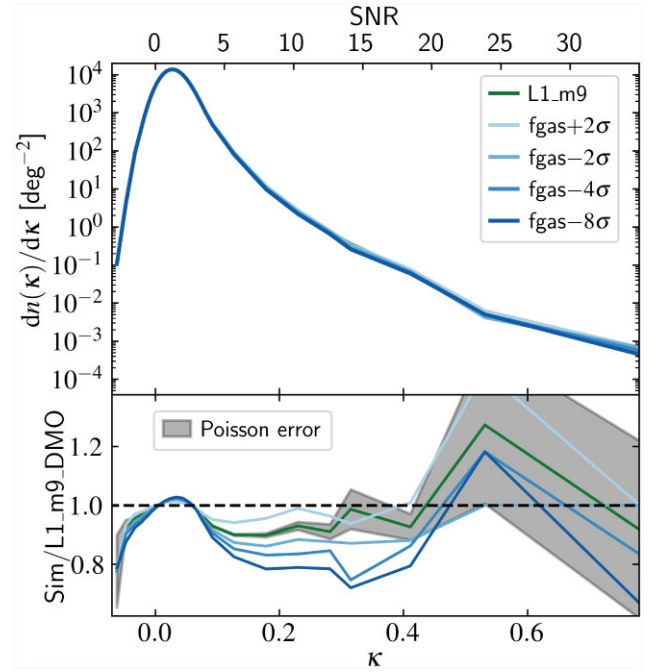


Figure 7. As Fig. 6 but comparing the runs calibrated to different gas fractions in cluster, as indicated by $f_{\text{gas}} - n\sigma$ where n is the change in the number of standard deviations compared with the fiducial L1_m9 (green curve) model, to L1_m9.DMO. Models with stronger feedback, indicated by an increasingly darker blue colour, have progressively stronger suppressed number densities of WL peaks.

The suppression can be understood by considering the impact of feedback, which expels gas from haloes and thus decreases their mass and thereby the WL potential.

4.4 Baryonic variations

Next, we compare the number density of peaks distributions in the variations that were calibrated to different observables. First, we compare the models that were calibrated to different gas fractions in clusters. We stress that these variations were run in similar boxes (L1) with the same resolution (m9), have identical subgrid feedback implementations, were run assuming the same cosmology, have the same initial conditions and the virtual observer was placed at the same position. The only difference between the runs is the value of four subgrid parameters, as described in Section 3.2, which were chosen to change the resulting gas fraction (f_{gas}) in clusters by a set amount. The number density of peaks for the runs with different gas fractions are shown in Fig. 7. Again, at $\kappa < 0.1$, the differences between the distributions are washed out due to the applied smoothing and noised.

In general, compared with the fiducial L1_m9, the lower (higher) gas fraction models correspond to simulations in which more (less) baryons have been evacuated from their haloes, as stronger (weaker) baryonic feedback is present. As the runs have the same initial conditions, the same haloes exist at the same positions in the simulation and comparisons between the runs are not affected by cosmic variance. We therefore include only the estimate of the L1_m9 Poisson noise in the lower panel. Any difference between the variations is a direct result of the halo mass differences as the WL kernel does not change. Fig. 7 shows that the models with lower gas fractions, which are indicated by an increasingly darker blue

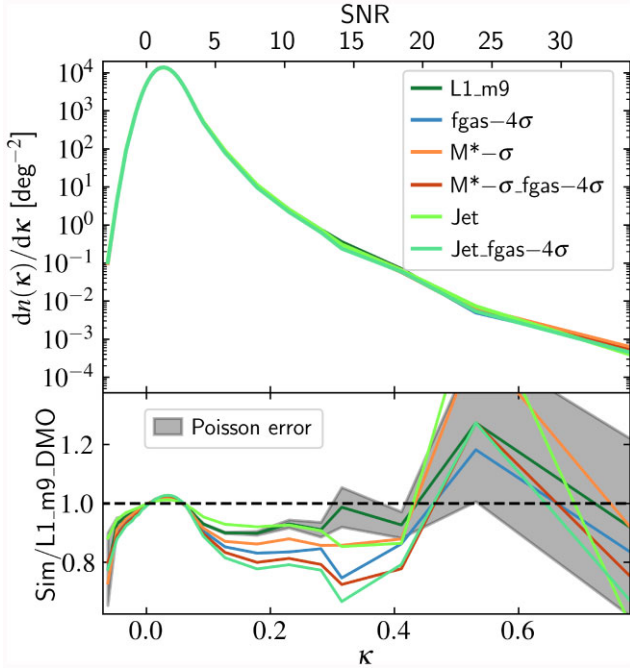


Figure 8. As Fig. 6 but comparing the baryonic feedback models varying the galaxy SMF (M^*), AGN feedback prescription (jet), gas fraction in clusters (f_{gas}), or a combination of these, to L1_m9_DMO. Calibration to lower galaxy SMFs leads to suppressed number densities of WL peaks. Variations in AGN subgrid prescriptions can lead to differences in WL peak counts not captured by the gas fraction in clusters.

colour, have progressively smaller number densities of WL peaks in the intermediate convergence regime ($\kappa \approx [0.1, 0.4]$), with the differences increasing up to $\kappa \approx 0.2$. We can understand the lower gas fractions as being the result of stronger feedback. Stronger feedback leads to more gas being expelled from the centres of haloes, and thus also to smaller overdensities. We therefore expect to see a stronger suppression of the WL peak counts for models calibrated to lower gas fractions. The hierarchy in gas fraction models agrees with the HMFs reported by Schaye et al. (2023), where they show that in the stronger feedback models, the HMF is increasingly more suppressed in the $10^{13} M_{\odot} \lesssim M_{200m} \lesssim 5 \times 10^{14} M_{\odot}$ regime. The number of haloes in this mass regime is smaller in the models with lower gas fractions and we thus expect there to be fewer high-valued WL convergence peaks too.

Next, we compare the remaining baryonic model implementations in Fig. 8. The L1_m9 (dark green) and jet (light green) models, which differ in the subgrid implementation of AGN feedback (thermal or jet) but have been calibrated to the same observables, show a different trend in their number density profiles. Whereas the baryonic suppression for L1_m9 is largest (10 per cent) at $\kappa = 0.1$, for model jet it increases to ≈ 20 per cent at $\kappa = 0.4$. In contrast, the $f_{\text{gas}}-4\sigma$ (blue) and $\text{jet}.f_{\text{gas}}-4\sigma$ (green) variations, which are each calibrated to the same observables but to a different gas fraction than the fiducial run, show a similar trend but a systematic difference of ≈ 10 per cent for $0.1 < \kappa < 0.4$, where the jet feedback model predicts smaller number densities. The difference between the two comparisons illustrates that the baryonic suppression of the number density of WL peaks cannot be expressed only in terms of the gas fraction in clusters, but also depends on how the astrophysical feedback is implemented and the gas distribution is reshaped, as jet feedback can potentially move mass further out (e.g. Federrath et al. 2014). In this case, we can

only partly understand the difference between the runs based on the HMFs. The HMFs of these runs in Schaye et al. (2023) show other differences. For halo masses of $10^{13} M_{\odot} \lesssim M_{200m} \lesssim 10^{14} M_{\odot}$, the thermal and jet models calibrated to the fiducial gas fraction show 5–20 per cent differences, with L1_m9 having a lower HMF, whereas the $f_{\text{gas}}-4\sigma$ models do not differ by more than 10 per cent from each other. At larger halo masses, up to $M_{200m} \approx 5 \times 10^{14} M_{\odot}$, the difference between L1_m9 and jet vanishes whereas the difference between the $f_{\text{gas}}-4\sigma$ and $\text{jet}.f_{\text{gas}}-4\sigma$ models remains of similar magnitude. Possibly, the peaks we measure primarily originate from haloes with $M_{200m} > 10^{14} M_{\odot}$, as the difference between the $f_{\text{gas}}-4\sigma$ models is larger in that regime. Liu & Haiman (2016) have used a halo model to study the origin of WL convergence peaks in the Canada–France–Hawaii Telescope Lensing Survey and found that the highest valued peaks are caused by a single massive halo of mass $M_{\text{vir}} \approx 10^{15} M_{\odot}$. It is unclear to what extent haloes of a certain mass contribute to which κ values of the WL peaks. We aim to investigate this in future research.

Next, we compare the models that have been calibrated to the same gas fraction but to different galaxy SMFs. Compared with the fiducial model, we see that the model with a lower galaxy SMF ($M^*-\sigma$; orange) also gives a suppressed number density of WL peaks. Within the simulation, to have a lower SMF, stronger feedback is required on galaxy scales. At the same time, the model was calibrated to have the same cluster gas fractions. The overall stronger feedback is reflected in the WL peak counts, as we see a suppression, suggesting the masses of the lenses are smaller than in the fiducial model. Compared with the difference between L1_m9 and $M^*-\sigma$, the difference between the $f_{\text{gas}}-4\sigma$ and $M^*-\sigma.f_{\text{gas}}-4\sigma$ models is slightly smaller but in general of similar magnitude and sign, suggesting the effects of galaxy- and cluster-scale feedback are (partly) separable. These differences are consistent with the HMF interpretation and results reported in Schaye et al. (2023). For $\kappa \gtrsim 0.4$, the regime dominated by Poisson noise and cosmic variance, the differences between the runs are less distinctive and informative.

Next, we consider the difference in the sensitivity to the two observables to which the FLAMINGO simulations were calibrated. Comparing the gas fraction variations in Fig. 7 with the stellar mass variations in Fig. 8, we see that $M^*-\sigma$ shows a similar suppression as the $f_{\text{gas}}-2\sigma$ model, both showing a suppression of ≈ 15 per cent for $\kappa = 0.1-0.4$, suggesting the WL observable is slightly more sensitive to the deviations as set by their current uncertainty in the SMF than the gas fraction in clusters. The $M^*-\sigma.f_{\text{gas}}-4\sigma$ model shows some additional suppression compared with $f_{\text{gas}}-4\sigma$ as the former run seems to fall in between $f_{\text{gas}}-4\sigma$ and $f_{\text{gas}}-8\sigma$, showing the suppression factors that are caused by lowering the two observables are at least partly independent. However, the suppression is smaller than that shown by the $\text{jet}.f_{\text{gas}}-4\sigma$ model, illustrating that different feedback prescriptions can already cause stronger differences than the shift in the SMF.

In summary, the suppression measured in the WL peak abundances between the FLAMINGO baryonic feedback variations is 5–25 per cent for WL convergence values of $\kappa = 0.1-0.4$, with models with stronger feedback showing a larger suppression. The peak abundance is sensitive to the amount of gas that is displaced from haloes and can generally be understood qualitatively by considering the impact of feedback on single haloes and the differences in the HMFs. The exact baryonic suppression of the WL peak counts is sensitive to both the subgrid feedback prescription, as varied between the thermal and jet models, as well as the feedback strength leading to different cluster gas fractions and galaxy SMFs.

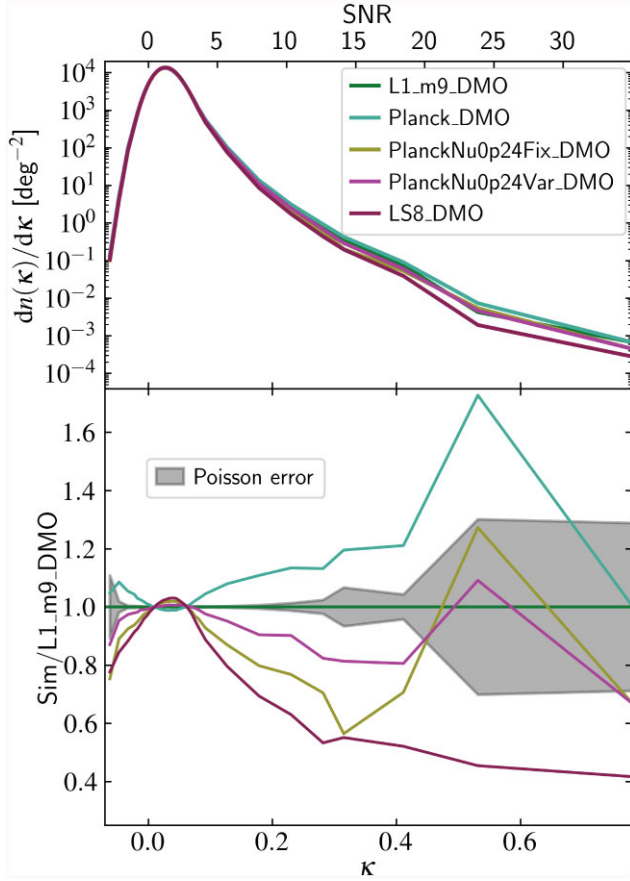


Figure 9. As Fig. 6 but comparing the cosmology variations in the 1 cGpc DMO box. The cosmologies are listed in Table 3. The vertical range in the bottom panel is larger than in all similar plots as the cosmology variations, in general, show larger differences than the baryonic feedback variations. The differences can be qualitatively understood from the impact of cosmological parameters (primarily Ω_m and σ_8) on the HMF.

4.5 Cosmology variations

In this section, we study the effect of varying the cosmology. The top panel in Fig. 9 shows the number density of the peaks measured for the different cosmology variations in the L1 box. The bottom panel again shows the ratio with the fiducial L1_m9 run. Note that the vertical range on the y-axis is larger than that of the previous plots, as the differences are, in general, larger. The most notably different model is the LS8 model. Across the entire κ range where the signal is not dominated by smoothing and noise, LS8 has a lower number density of peaks than any of the other models. Here, we can understand the differences between the different distributions by considering the impact of different cosmological parameters on the HMF. The HMF depends strongly on the cosmological parameters Ω_m and σ_8 , where a lower Ω_m decreases the overall amplitude of the HMF and σ_8 primarily influences the high mass end of the HMF, moving the exponential cutoff to lower masses for lower values of σ_8 (e.g. Tinker et al. 2008; Xhakaj et al. 2023). Compared with the fiducial model, the LS8 model has almost the same value of Ω_m but a 6 per cent lower value of σ_8 . We thus expect fewer massive haloes in the LS8 model. This is reflected in Fig. 9, where we see that the number density of WL peaks is the lowest for the LS8 run across the entire intermediate convergence regime ($\kappa \approx [0.1, 0.4]$). As shown by Li et al. (2019), for a Stage IV WL survey, WL peaks

with $\text{SNR} > 3$ each originate from a single massive halo along the line of sight, and the number of these haloes is thus affected by the lower σ_8 value. Their SNR value does not directly correspond to our $\text{SNR} = 3$ as the noise, smoothing, and redshift distribution are different. Nevertheless, we understand the highest peaks to be caused by a single halo. We therefore expect the number density of (high) WL peaks to be lower in the LS8 run. The deviation from the fiducial model increases for larger convergence values. Following similar reasoning, we can understand why the Planck cosmology model shows the largest number density of peaks. Compared with the fiducial cosmology, the Planck cosmology has larger values of both Ω_m and σ_8 . We thus expect the total number of haloes to be larger as well as the cutoff at high halo mass to shift to higher masses, which is reflected in the bottom panel of Fig. 9, as the Planck model has larger number densities than the other variations for $\kappa > 0.1$.

Our results agree qualitatively with Coulton et al. (2020), who varied the cosmological parameters Ω_m and A_s . Our analysis, which extends to larger κ values, shows that the model with a higher total matter density continues to have an enhanced number density of WL peaks for the entire WL convergence regime. In our case, in the intermediate convergence regime ($\kappa \approx [0.1, 0.4]$), the difference increases from 10 to 20 per cent. In both instances, the model with the lower value of A_s , in our case the LS8 model, has a slightly increased number density of WL peaks for $\text{SNR} = 0-3$, whereas for larger values, the runs with the lower values of A_s show a suppression of the number density of WL peaks. Our analysis shows this holds up to the largest values of κ .

We now turn to the models with varying neutrino masses. We observe that the number density of peaks corresponding to the heavier neutrino models (Planck0p24Fix and Planck0p24Var), in the intermediate WL convergence regime, is 20–40 per cent lower compared with the Planck model. In cosmology, we understand neutrinos to act as a form of hot DM. Due to their high speeds and weak interactions with regular matter, neutrinos cannot be contained effectively in regions smaller than their free-streaming length. Therefore, they can carry mass away from overdense regions, impeding the growth of clusters (e.g. Lesgourgues & Pastor 2006). The effect of neutrinos on the HMF has been studied in the BAHAMAS simulation by Mummery et al. (2017), who showed that the massive end of the HMF is preferentially suppressed by the free streaming of massive neutrinos. In their comparison, the difference in the HMF in the halo mass regime of $[10^{14}-10^{15}] M_\odot$ is 10–20 per cent for a similar neutrino mass difference as we consider. Model Planck0p24Fix, for which the values of the cosmological parameters other than the neutrino mass are the same as for model Planck, shows a 10 per cent stronger suppression than the Planck0p24Var model. Qualitatively, we find similar results to Coulton et al. (2020), who studied the peak count distribution with varying cosmological parameters in the N -body MASSIVENUS simulations (Liu et al. 2018), as we see that more massive neutrinos suppress the WL peak counts. Quantitatively, our suppression is a few factors larger than they report. The most obvious explanation is that we have $\Delta \sum m_\nu c^2 = 0.18$ eV, whereas they have a summed mass difference of 0.1 eV. We thus expect the suppression we measure to be larger than the value they find. Our results also agree quantitatively with Ferlito et al. (2023) who studied peak counts in MTNG variations with varying neutrino mass. They consider two MTNG variations with $\sum m_\nu = 0.1$ and 0.3 eV, whose summed difference is 0.02 eV larger than in our analysis (we compare $\sum m_\nu = 0.06-0.24$). They too find a suppression of 20–30 per cent in the $\kappa = 0.1-0.2$ domain. Finally, Fong et al. (2019) studied the impact of massive neutrinos on WL peak counts in the BAHAMAS simulation. Amongst others, they consider models with $\sum m_\nu = 0.06$

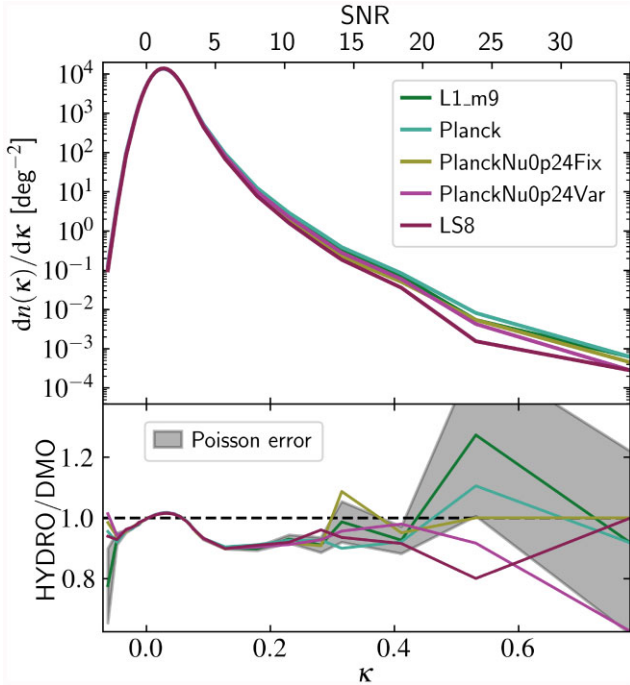


Figure 10. As Fig. 6 but comparing the hydrodynamical runs with different cosmologies but the same baryonic calibration with their corresponding DMO runs. Whereas the variation between the DMO signals in Fig. 9 is up to 300 per cent, the baryonic suppression (as seen in the bottom panel) is similar for all cosmology variations, illustrating the separability of the cosmology and baryonic impact.

and 0.24 eV. They see a difference of 10–20 per cent up to $\text{SNR} = 9$, slightly smaller than the difference we find.

The comparisons in this section show that the number densities of peaks are highly sensitive to cosmology. Variations of a few per cent in Ω_m or σ_8 cause changes in the WL peak abundances of ≈ 20 –60 per cent. The differences can be understood qualitatively by considering the impact of cosmological parameters on the HMF.

4.6 Separability of cosmological and hydrodynamical effects

Finally, we explore the separability of the cosmological and astrophysical impact on the number density of WL peaks. In Fig. 10, the top panel shows the number density of WL peaks for the cosmology variations, which were done at fixed baryonic calibration, in the full-hydrodynamical runs. The bottom panel shows, for each cosmology variation separately, the ratio to its corresponding DMO run. Despite the big differences seen between the cosmology variations in Fig. 9, when comparing the variations with their own DMO run, the baryonic suppression is very similar for all cosmology variations, as shown in the bottom panel of Fig. 10, illustrating that the cosmological and hydrodynamical effect are largely separable. The deviation between the number densities of WL peaks up to $\kappa = 0.2(0.4)$ is not more than 1(10) per cent, and generally falls within the Poisson noise for L1_m9 as indicated by the grey-shaded area, whereas the cosmology variations in Fig. 9 show up to an order of magnitude larger differences in the same κ regime.

We can understand the separability from the different impacts of the two processes on haloes. We have seen that the cosmology dependence translates itself into changes in the total number of haloes at fixed mass (through Ω_m and σ_8) and their halo mass (through mass being carried away by neutrinos), whereas the baryonic physics

displaces gas and removes mass from the centres of haloes. As these processes are physically independent we expect to see the cosmology and baryonic separability for the number density of WL peaks. The comparison shows that the two processes are indeed separable, which demonstrates the potential to model the impact of cosmology on WL peaks based on DMO simulations only.

5. DISCUSSION

We start the discussion by comparing the different sets of baryonic and cosmological variations of the previous sections. First, we compare the impact of all baryonic variations to the difference in suppression between the fiducial DMO and hydrodynamical simulation. Fig. 6 shows that the fiducial baryonic suppression is ≈ 10 per cent for $\kappa = 0.1$ and that the suppression decreases for larger WL convergence values. Comparing this with the suppression and enhancement of the different feedback variations in Figs 7 and 8, we see that the other baryonic feedback variations show a suppression of 5–30 per cent compared with the DMO signal, for $\kappa = [0.1, 0.4]$.

Comparing the impact of cosmology and baryonic physics, we can study the degeneracy of the two effects. With the limited number of variations that we consider in this research, we do not find clear evidence that the two effects are non-degenerate, meaning that the two effects impact the same κ ranges similarly, and the two processes each have to be modelled carefully. However, we stress that we only have a limited number of variations and that to properly address this question one should compare variations that jointly vary in cosmology and baryonic physics, which we aim to do in the future. Additionally, including other observables may help break the degeneracy.

The variations between the fiducial and the Planck or LS8 models, as shown in Fig. 9 are larger than the differences between the runs that vary the gas fractions, as seen in Fig. 7, which span 10σ in gas fractions. The enhancement in the number density of the Planck cosmology is about twice as large as for the $f_{\text{gas}} + 2\sigma$ model and LS8 suppresses the number density of peaks 2–3 times as much as the $f_{\text{gas}} - 8\sigma$ model. These differences illustrate that within the FLAMINGO simulation suite, the number density of peaks is more sensitive to reasonable differences in cosmological parameters than to the broad range of explored gas fractions, some of which are already ruled out by present observations of galaxy clusters (assuming there are no unrecognized systematics). It is important to keep in mind that even among models that reproduce the observed galaxy SMF and predict the same cluster gas fractions, the exact effect of the baryonic feedback can depend on the adopted subgrid model, as we saw in the comparison of the jet to thermal AGN feedback models. Other parts of the astrophysics model could also play an important role. However, since the covered range in gas fractions is large, the analysis still highlights that the number density of peaks is more sensitive to the total amount of matter, set by Ω_m , and its clumpiness, as set by σ_8 , than to the uncertainty in the distribution of gas in the Universe, resulting from baryonic feedback. Nevertheless, the astrophysical uncertainties are not negligible compared with the effect of varying the cosmology.

Next, we focus on the different impacts of neutrinos and baryons. To this end, we compare the relative difference of the gas fraction models with the fiducial model (Fig. 7) and the heavier neutrino models to the Planck model (Fig. 9), as the heavier neutrino models are variations on the latter model. The suppression by the neutrino models with $\sum m_\nu = 0.24$ eV in the intermediate convergence regime ($\kappa \approx [0.1, 0.4]$) is slightly stronger than the difference between $f_{\text{gas}} + 2\sigma$ and $f_{\text{gas}} - 8\sigma$. Our analysis thus shows that WL peak counts

are more sensitive to the difference in cosmological parameters between the two sets of cosmologies currently favoured by either CMB or WL measurements than the uncertainties in the impact of SN and AGN feedback leading to different cluster gas fractions or galaxy SMFs.

While the cosmology variations cause larger differences in the number density of peaks, it is clear that the uncertainty on WL peaks arising from baryonic effects is not negligible for a Stage IV inference. Both the cosmology and baryonic variations considered in this paper are on the extreme side in the sense that we expect Stage IV WL surveys to be able to discriminate between them, which means that the baryons have to be properly modelled. In particular, the impact of baryons and cosmology shows up in the same convergence regime with similar behaviour, but Section 4.6 has illustrated that the cosmology and baryonic impact are mostly separable in the sense that the baryonic suppression is insensitive to the cosmology. The uncertainties due to baryons may be overcome by considering a joint inference of two-point and non-Gaussian statistics. For example, Semboloni et al. (2013) have shown that baryonic feedback impacts two- and three-point shear statistics differently. Using a combination of different WL statistics, the uncertainties due to baryons could be calibrated simultaneously whilst constraining the cosmology.

Finally, although the low- κ regime ($\kappa < 0.1$) is strongly affected by smoothing and noise, there are still clear differences between the simulations. For instance, whereas LS8 has a suppressed number density of WL peaks at the intermediate- and high- κ end, it has an increased number of peaks around $\kappa = 0.05$ (Fig. 9). Although the difference is only a few per cent, this regime contains several orders of magnitude more peaks, and therefore the impact of Poisson noise and cosmic variance (Fig. 6) vanishes, and it may potentially help in discriminating between the models. However, the interpretation of these peaks is more difficult because the systematics affecting this regime need to be very well understood. Also, the physical interpretation of these peaks is less straightforward as they are caused by multiple objects along the line of sight (Yang et al. 2011), and we note that the baryonic impact and cosmology variations are of similar magnitude around $\kappa = 0.05$, whereas the variations are relatively greater for the cosmology variations in the larger κ regime. We aim to explore both the constraining power and the physical origin of these peaks in future research.

6. SUMMARY AND CONCLUSIONS

We have studied the impact of baryonic feedback on the distribution of peaks in WL convergence maps. We used the state-of-the-art cosmological hydrodynamical FLAMINGO simulation suite, in which the parameters of the subgrid prescription for feedback are calibrated to match the observed $z = 0$ galaxy SMF and the gas fraction in low- z clusters. To mimic the signal of a Stage IV WL survey, we used a full-sky ray-tracing method with a *Euclid*-like source redshift distribution and shape measurement noise. We studied numerical convergence (Fig. 4) and cosmic variance (Fig. 5), showing that both are under control in the intermediate WL convergence regime ($\kappa \approx [0.1, 0.4]$), where most cosmological information is contained. Our analysis shows that for $\kappa \gtrsim 0.4$, the signal is dominated by Poisson noise and cosmic variance, which indicates a limit to the usefulness of high- κ WL peaks as all WL surveys will suffer from these effects.

Our results agree with previous studies (Coulton et al. 2020; Osato et al. 2021; Ferlito et al. 2023) as we find that baryonic feedback processes, which expel gas from the centres of haloes and thereby decrease their mass, lead to a suppression of the number density of

WL peaks of ~ 10 per cent in the cosmologically informative WL convergence regime (Fig. 6).

At fixed cosmology, initial conditions, and observer position, we compared nine models that differ in their galaxy SMF, cluster gas fraction, and/or AGN feedback implementation; we find baryonic suppressions in the range of 5–30 per cent (Figs 7 and 8). Our results show a clear trend with cluster gas fraction; in simulations with a lower gas fraction, the number density of peaks is more suppressed. The differences may be understood from an individual halo perspective: Cluster haloes with lower gas fractions have lost more gas from their centre, resulting in a lower mass and hence a suppression of the WL signal and number density of peaks. Similarly, calibrating the feedback prescription so as to reduce the amplitude of the galaxy SMF also lowers the number density of WL peaks in the well-resolved convergence regime ($\kappa \approx [0.1, 0.4]$). The differences between the thermal and kinetic jet AGN feedback models, calibrated to obtain the same galaxy mass function and cluster gas fractions, illustrate that the baryonic suppression is not fully specified by these two observables but also depends on the details of the feedback implementation.

We compared simulations with different values of the cosmological parameters, favoured either by the CMB or by WL measurements, and with different neutrino masses. The impact of these variations can be understood by reference to the effect that the cosmological parameters have on the HMF. The cosmology variations show greater differences than the baryonic variations at fixed cosmology, suppressing (enhancing) the number density of WL peaks up to 60(20) per cent (Fig. 9).

Our results illustrate that WL peak counts are a useful statistic to constrain cosmological parameters with upcoming WL surveys. Within the FLAMINGO simulation suite, we have shown that variations in cosmology have a larger impact than variations in the treatment of baryonic physics. However, the differences induced by baryons are larger than the expected accuracy of upcoming surveys and must therefore be modelled carefully. Importantly, our analysis shows that while the effects of baryonic physics are well behaved and can be understood, their effects are qualitatively similar to those due to variations in cosmology. However, the impact of baryons is insensitive to the changes in cosmology that we explored, suggesting that the effect of cosmology can be investigated using DMO simulations.² The analysis shows that the deviation of baryonic suppression at different cosmologies up to $\kappa = 0.2(0.4)$ does not exceed 1(10) per cent (Fig. 10) and that the $\kappa > 0.4$ regime is dominated by Poisson noise and cosmic variance.

Non-Gaussian statistics could be used to disentangle baryonic from cosmological effects. In order to exploit WL peaks or other beyond-Gaussian statistics, both neutrinos and baryons need to be carefully modelled. Multiple WL statistics could be used to simultaneously calibrate the baryonic physics and to constrain the cosmology. Additionally, it will be necessary to forward model photometric redshifts, intrinsic alignments, masks, multiplicative shear bias as well as tomographical analyses (e.g. Fluri et al. 2018; Zürcher et al. 2021; Zhang et al. 2022). Ideally, the simulations will cover a broad enough range of cosmological and astrophysical feedback parameters. In the future, such approaches will become possible by using emulators built on top of simulation suites in which the feedback implementation and the cosmological parameters are varied simultaneously.

²Note in particular that large variations in Ω_b/Ω_m have not been explored in this study.

ACKNOWLEDGEMENTS

JCB thanks Christopher Davies for the helpful discussions on peaks in HEALPix maps. CSF acknowledges support by the European Research Council (ERC) through Advanced Investigator grant, DMIDAS (GA 786910). This work is partly funded by research programme Athena 184.034.002 from the Dutch Research Council (NWO). This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (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 data supporting the plots within this article are available on reasonable request to the corresponding author. The FLAMINGO simulation data will eventually be made publicly available.

REFERENCES

- Abbott T. M. C. et al., 2019, *ApJ*, 872, L30
- Abbott T. M. C. et al., 2022, *Phys. Rev. D*, 105, 023520
- Abdalla E. et al., 2022, *J. High Energy Astrophys.*, 34, 49
- Akino D. et al., 2022, *PASJ*, 74, 175
- Amon A. et al., 2022, *Phys. Rev. D*, 105, 023514
- Amon A. et al., 2023, *MNRAS*, 518, 477
- Anderson L. et al., 2014, *MNRAS*, 441, 24
- Aricò G., Angulo R. E., Hernández-Monteagudo C., Contreras S., Zennaro M., Pellejero-Ibañez M., Rosas-Guevara Y., 2020, *MNRAS*, 495, 4800
- Asgari M. et al., 2021, *A&A*, 645, A104
- Asgari M., Mead A. J., Heymans C., 2023, *Open J. Astrophys.*, 6, 39
- Barreira A., Bose S., Li B., Llinares C., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 031
- Bartelmann M., Schneider P., 2001, *Phys. Rep.*, 340, 291
- Becker M. R., 2013, *MNRAS*, 435, 115
- Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143
- Bernardeau F., van Waerbeke L., Mellier Y., 1997, *A&A*, 322, 1
- Blas D., Lesgourgues J., Tram T., 2011, *J. Cosmol. Astropart. Phys.*, 2011, 034
- Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, *MNRAS*, 456, 2361
- Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
- Borrow J., Schaller M., Bower R. G., Schaye J., 2022, *MNRAS*, 511, 2367
- Bradač M., Schneider P., Lombardi M., Erben T., 2005, *A&A*, 437, 39
- Chaikin E., Schaye J., Schaller M., Benítez-Llambay A., Nobels F. S. J., Ploekinger S., 2023, *MNRAS*, 523, 3709
- Chisari N. E. et al., 2018, *MNRAS*, 480, 3962
- Clark S. J., Vattis K., Fan J., Koushiappas S. M., 2023, *Phys. Rev. D*, 107, 083527
- Coulton W. R., Liu J., McCarthy I. G., Osato K., 2020, *MNRAS*, 495, 2531
- Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, *MNRAS*, 423, 2279
- Dalla Vecchia C., Schaye J., 2008, *MNRAS*, 387, 1431
- Davies C. T., Cautun M., Giblin B., Li B., Harnois-Déraps J., Cai Y.-C., 2021, *MNRAS*, 507, 2267
- Davies C. T., Cautun M., Giblin B., Li B., Harnois-Déraps J., Cai Y.-C., 2022, *MNRAS*, 513, 4729
- Debackere S. N. B., Hoekstra H., Schaye J., 2022, *MNRAS*, 515, 6023
- Dietrich J. P., Hartlap J., 2010, *MNRAS*, 402, 1049
- Dodelson S., Zhang P., 2005, *Phys. Rev. D*, 72, 083001
- Elbers W., Frenk C. S., Jenkins A., Li B., Pascoli S., 2021, *MNRAS*, 507, 2614
- Elbers W., Frenk C. S., Jenkins A., Li B., Pascoli S., 2022, *MNRAS*, 516, 3821
- Euclid Collaboration, 2019, *A&A*, 627, A59
- Euclid Collaboration, 2020, *A&A*, 642, A191
- Euclid Collaboration, 2023, *A&A*, 675, A120
- Fabbian G., Calabrese M., Carbone C., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 050
- Federrath C., Schrön M., Banerjee R., Klessen R. S., 2014, *ApJ*, 790, 128
- Feldbrugge J., van Engelen M., van de Weygaert R., Pranav P., Vegter G., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 052
- Ferlito F. et al., 2023, *MNRAS*, 524, 5591
- Fluri J., Kacprzak T., Sgier R., Refregier A., Amara A., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 051
- Fong M., Choi M., Catlett V., Lee B., Peel A., Bowyer R., King L. J., McCarthy I. G., 2019, *MNRAS*, 488, 3340
- Giocoli C. et al., 2016, *MNRAS*, 461, 209
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
- Gouin C. et al., 2019, *A&A*, 626, A72
- Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, *Astrophysics Source Code Library*, record ascl:2008.024
- Hamana T. et al., 2020, *PASJ*, 72, 16
- Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, *A&A*, 499, 31
- Hilbert S. et al., 2020, *MNRAS*, 493, 305
- Hildebrandt H. et al., 2017, *MNRAS*, 465, 1454
- Hoekstra H., Jain B., 2008, *Annu. Rev. Nucl. Part. Sci.*, 58, 99
- Hu W., 2000, *Phys. Rev. D*, 62, 043007
- Huškø F., Lacey C. G., Schaye J., Schaller M., Nobels F. S. J., 2022, *MNRAS*, 516, 3750
- Jain B., Seljak U., White S., 2000, *ApJ*, 530, 547
- Kaiser N., 1986, *MNRAS*, 222, 323
- Kaiser N., 1992, *ApJ*, 388, 272
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- Kilbinger M., 2015, *Rep. Prog. Phys.*, 78, 086901
- Kratochvil J. M., Haiman Z., May M., 2010, *Phys. Rev. D*, 81, 043519
- Kratochvil J. M., Lim E. A., Wang S., Haiman Z., May M., Huppenberger K., 2012, *Phys. Rev. D*, 85, 103513
- Kugel R. et al., 2023, *MNRAS*, 526, 6103
- Kuijken K. et al., 2019, *A&A*, 625, A2
- Lahav O., Liddle A. R., 2022, preprint ([arXiv:2201.08666](https://arxiv.org/abs/2201.08666))
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Lee M. E., Lu T., Haiman Z., Liu J., Osato K., 2023, *MNRAS*, 519, 573
- Lesgourgues J., Pastor S., 2006, *Phys. Rep.*, 429, 307
- Li Z., Liu J., Zorrilla Matilla J. M., Coulton W. R., 2019, *Phys. Rev. D*, 99, 063527
- Limber D. N., 1953, *ApJ*, 117, 134
- Liu J., Haiman Z., 2016, *Phys. Rev. D*, 94, 043533
- Liu J., Petri A., Haiman Z., Hui L., Kratochvil J. M., May M., 2015, *Phys. Rev. D*, 91, 063507
- Liu J., Bird S., Zorrilla Matilla J. M., Hill J. C., Haiman Z., Madhavacheril M. S., Petri A., Spergel D. N., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 049
- LoVerde M., Afshordi N., 2008, *Phys. Rev. D*, 78, 123506
- LSST Science Collaboration, 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- Lu T., Haiman Z., 2021, *MNRAS*, 506, 3406
- McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, *MNRAS*, 465, 2936
- McCarthy I. G., Bird S., Schaye J., Harnois-Déraps J., Font A. S., van Waerbeke L., 2018, *MNRAS*, 476, 2999
- McCarthy I. G. et al., 2023, *MNRAS*, 526, 5494
- McClintock T. et al., 2019, *ApJ*, 872, 53
- Marques G. A. et al., 2024, *MNRAS*, 528, 4513
- Michaux M., Hahn O., Rampf C., Angulo R. E., 2021, *MNRAS*, 500, 663
- Muciaccia P. F., Natoli P., Vittorio N., 1997, *ApJ*, 488, L63
- Mummery B. O., McCarthy I. G., Bird S., Schaye J., 2017, *MNRAS*, 471, 227
- Osato K., Liu J., Haiman Z., 2021, *MNRAS*, 502, 5593
- Pakmor R. et al., 2023, *MNRAS*, 524, 2539

- Petri A., Haiman Z., Hui L., May M., Kratochvil J. M., 2013, *Phys. Rev. D*, 88, 123002
- Petri A., Haiman Z., May M., 2017, *Phys. Rev. D*, 95, 123503
- Planck Collaboration VI, 2020, *A&A*, 641, A6
- Ploechinger S., Schaye J., 2020, *MNRAS*, 497, 4857
- Pratten G., Lewis A., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 047
- Price M. A., McEwen J. D., Pratley L., Kitching T. D., 2021, *MNRAS*, 500, 5436
- Sabyr A., Haiman Z., Matilla J. M. Z., Lu T., 2022, *Phys. Rev. D*, 105, 023505
- Salcido J., McCarthy I. G., Kwan J., Upadhye A., Font A. S., 2023, *MNRAS*, 523, 2247
- Schaller M. et al., 2023, preprint (arXiv:2305.13380)
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, 383, 1210
- Schaye J. et al., 2023, *MNRAS*, 526, 4978
- Schneider P., 2016, *A&A*, 592, L6
- Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 020
- Schneider A., Stoira N., Refregier A., Weiss A. J., Knabenhans M., Stadel J., Teyssier R., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 019
- Schöneberg N., Abellán G. F., Sánchez A. P., Witte S. J., Poulin V., Lesgourgues J., 2022, *Phys. Rep.*, 984, 1
- Schrabback T. et al., 2018, *MNRAS*, 474, 2635
- Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, *MNRAS*, 417, 2020
- Semboloni E., Hoekstra H., Schaye J., 2013, *MNRAS*, 434, 148
- Shirasaki M., Hamana T., Yoshida N., 2015, *MNRAS*, 453, 3043
- Simon P., Hilbert S., 2018, *A&A*, 613, A15
- Spergel D. et al., 2015, preprint (arXiv:1503.03757)
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Stanek R., Rudd D., Evrard A. E., 2009, *MNRAS*, 394, L11
- Takahashi R., Nishimichi T., Namikawa T., Taruya A., Kayo I., Osato K., Kobayashi Y., Shirasaki M., 2020, *ApJ*, 895, 113
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
- Upadhye A. et al., 2023, preprint (arXiv:2308.09755)
- Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, *MNRAS*, 442, 2641
- Weiss A. J., Schneider A., Sgier R., Kacprzak T., Amara A., Refregier A., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 011
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, *MNRAS*, 399, 574
- Xhakaj E., Leauthaud A., Lange J., Krause E., Hearin A., Huang S., Wechsler R. H., Heydenreich S., 2023, preprint (arXiv:2306.03777)
- Yang X., Kratochvil J. M., Wang S., Lim E. A., Haiman Z., May M., 2011, *Phys. Rev. D*, 84, 043529
- Yang X., Kratochvil J. M., Huppenberger K., Haiman Z., May M., 2013, *Phys. Rev. D*, 87, 023511
- Zhang T., Liu X., Wei C., Li G., Luo Y., Kang X., Fan Z., 2022, *ApJ*, 940, 96
- Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *J. Open Source Softw.*, 4, 1298
- Zorrilla Matilla J. M., Waterval S., Haiman Z., 2020, *AJ*, 159, 284
- Zürcher D., Fluri J., Sgier R., Kacprzak T., Refregier A., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 028
- van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, *MNRAS*, 415, 3649
- van Daalen M. P., McCarthy I. G., Schaye J., 2020, *MNRAS*, 491, 2424

APPENDIX A: BOX ROTATION

In this appendix, we explore the impact of different box rotation strategies. As all the baryonic feedback variations were carried out in 1 cGpc boxes, which need to be replicated six times to cover the distance to $z = 3$, the approximate redshift range necessary for a Stage IV WL survey (see Fig. 1 and the green arrow indicating the L1 box length). As the light-cone extends to $z = 3$ in all directions, in total, 13 L1 box replications are necessary for the L1 runs.

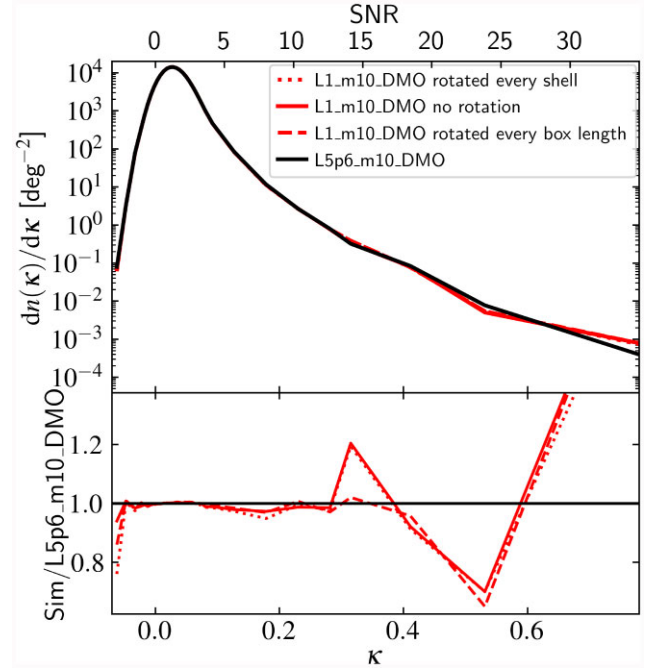


Figure A1. Top panel: Number densities of WL peaks for the L5p6.m10.DMO model (black) and L1.m10.DMO run (red) with different mass shell rotation strategies where the shells in the latter simulation are not rotated (dashed), every shell is randomly rotated (dotted), or the shells are randomly rotated every half box length (solid). Bottom panel: Ratio to the L5p6.m10.DMO signal. The different rotation methods only show minor differences.

To test the impact of the replication along the line of sight and the strategy to deal with the shell rotations, we compare the signal of the L1.m10.DMO run to that of the L5p6.m10.DMO run. This simulation, which contains 5040^3 DM and 2800^3 neutrino particles, was run in a 5.6 Gpc box at the fiducial cosmology and has the same resolution as the low-resolution (m10) L1 run. For this cosmology, 5.6 cGpc corresponds to a redshift of $z = 2.2$, which covers the largest part of the redshift range as indicated in Fig. 1. However, as the light-cone extends in all directions, the independent volume extends only to $L = 2.8$ cGpc ($z = 0.8$), which corresponds approximately to the mean of the redshift distribution we use (see Fig. 1). Here, to make sure we do not repeat over any structures in the 5.6 cGpc box, we adopt a delta-function source redshift distribution, $n(z) = \delta(z - 0.8)$.

We then compare L5p6.m10.DMO with L1.m10.DMO for different treatments of the redshift shells, where we consider the signal from the 5.6 cGpc run to be the ground truth. For L1.m10.DMO, we test three different rotation prescriptions. First, we apply no rotation to any of the mass shells. In this case, depending on the angular position on the sky, we sum up to six times over the same structure (though at different stages of evolution) along the line of sight. Second, we rotate every shell randomly to prevent running into the same structure. Third, we randomly rotate the shells whenever the diameter of the light-cone is larger than the box length. In this way, we still minimize the number of repetitions but also do not superfluously erase any possible correlations along the line of sight.

Fig. A1 shows the number densities of peaks for the different methodologies. The L1.m10.DMO variants are indicated in red where the no rotation, rotation of every shell, and rotation of every half box length are indicated by the solid, dotted, and dashed lines, respectively. The bottom panel compares the three rotation

strategies with the signal measured in the L5p6_m10_DMO run, which corresponds to the black curve in the top panel. As the different box sizes are different realizations of the same Universe, we do not expect them to show perfect agreement but be impacted by cosmic variance. As is clear from the bottom panel, the variations in replication strategy have only a minor impact on the measured signal. This agrees with Upadhye et al. (2023), who studied similar rotation strategies for the CMB lensing angular power spectrum. All three rotations show some minor suppression compared with the L5p6 box, independent of the rotation strategy. Only at $\kappa \approx 0.3$, there is a difference between the rotation strategies. Here, the not rotated and rotated every shell procedures give a 20 per cent difference while the rotated every half box length method yields almost perfect agreement with L5p6_m10_DMO. In our analysis, we choose to randomly rotate every half box length and we apply the same rotation to simulations of the same box size.

APPENDIX B: SMOOTHING AND NOISE

In this appendix, we illustrate the effect of the applied galaxy shape noise and smoothing on the number densities of WL convergence peaks. The noise and smoothing are applied to mimic the expected observed galaxy shape noise (equation 15) and the effect of any instrument carrying out a WL survey. We compare distributions without noise and smoothing to those with noise and smoothing in Fig. B1. If smoothing is applied, the map is smoothed with a Gaussian kernel with FWHM = 1 arcmin. The figure includes the mean of the eight virtual observers in the hydrodynamical L2p8 (red) and L2p8_DMO (black) runs. The solid and dotted lines correspond to instances in which the final convergence map contains no smoothing and noise, or both noise and smoothing, respectively. The red curves are directly behind their corresponding black curves.

The bottom panel in Fig. B1 gives the ratio of the hydrodynamical to DMO number densities as measured for the maps that used the same noise and smoothing treatment. Depending on the convergence regime, the noise and smoothing have a different impact. For $\kappa \approx 0$, the smoothing and noise wash out the largest baryonic effects, as can be seen in the bottom panel where the difference at $\kappa = 0$ is suppressed from 10 to 2 per cent. The ratio of the maps without any treatment shows an enhancement of peaks with $\kappa \approx 0$ in the hydrodynamical run, which has almost completely vanished after the smoothing and noise are applied. For $\kappa > 0.1$, the number of peaks is suppressed significantly by the map treatment, as can be seen by comparing the dashed with solid lines in the top panel. The smoothing primarily impacts the total number of peaks and the

high- κ end as it levels out fluctuations on the smallest scales and lowers the values of the highest peaks, whereas the noise is primarily responsible for erasing the difference between the runs around $\kappa = 0$. The comparison illustrates that the entire κ regime is sensitive to the treatment of the WL convergence maps, which makes sense as smoothing the maps effectively smooths out overdensities, but it does illustrate that when comparing with observations a more dedicated study into these systematics should be carried out.

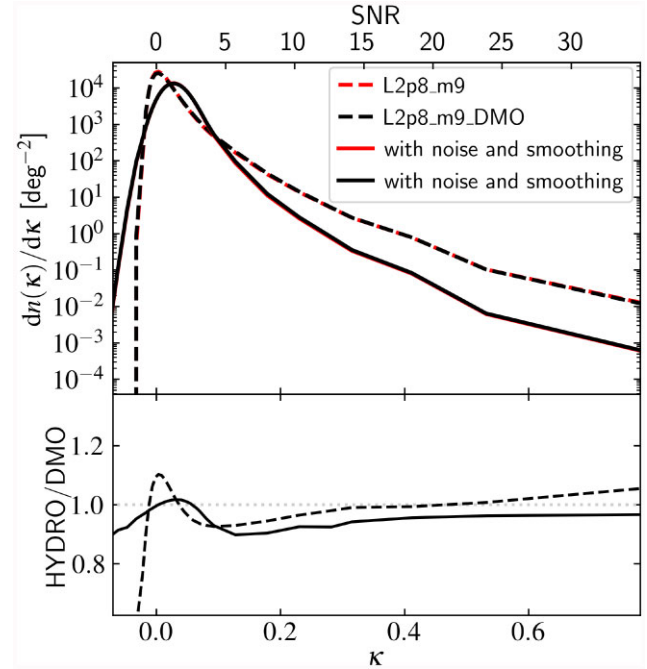


Figure B1. Top panel: Number densities of WL peaks for the mean of the eight observers in the L2p8_m9 (red) and L2p8_m9_DMO (black) runs with different treatment for the applied smoothing and noise. The solid and dashed lines correspond to number densities measured from maps that have no smoothing or noise, or both noise and smoothing, respectively. The lines with the same linestyle are directly behind each other. The bottom panel shows the ratio of the hydrodynamical to DMO number densities for maps that have had the same noise and smoothing treatment. The entire WL convergence regime is impacted by the smoothing and noise, indicating a careful analysis has to be done when carrying out an inference with real observations.

This paper has been typeset from a \LaTeX file prepared by the author.