A Peak into the Power of the Universe

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr. S. de Rijcke,
volgens besluit van het college voor promoties
te verdedigen op donderdag 21 januari 2027
klokke 11:30 uur

door

Jeger Cornelis Broxterman
geboren te Nieuwegein, Nederland
in 1999

Promotores:
Prof.dr. H. Hoekstra
Prof.dr. J. Schaye

Co-promotor:
Dr. M. Schaller

Promotiecommissie:
Dr. A. Amon (Princeton University)
Prof.dr. R.E. Angulo (Donostia International Physics Center)
Prof.dr. K.H. Kuijken
Prof.dr. A. Silvestri
Prof.dr. I.A.G. Snellen

Copyright © *2026 Jeger Broxterman*

ISBN: 9789464966299

Cover: artist's impression of weak gravitational lensing. The paths of photons as seen by the observer are bent due to matter curving spacetime. The spherical structure represents the FLAMINGO light-cones, used for weak lensing in simulations in this thesis. The galaxy is *Euclid*'s view of the spiral galaxy IC 342, symbolising the observations used in this work. Design by Florian Tripoteau.



Universiteit Leiden

Printed by Gildeprint

'Eerlijk gezegd,' zei Max, 'als jongen al heb ik nooit begrepen, dat iemand zich klein kon voelen tegenover het heelal. De mens weet *toch hoe overweldigend groot het is, en nog een paar dingen meer ook, – dan is hij toch niet klein! Dat hij dat allemaal heeft ontdekt, bewijst toch juist zijn grootheid. Het verbazende is eerder, dat dat nietige wezen het hele universum in die minieme ruimte onder zijn schedeldak kan bevatten, – en daar bovendien nog over kan reflecteren, zoals wij nu doen. Dat maakt hem in zekere zin zelfs* groter *dan het heelal.'*
- Harry Mulisch, De ontdekking van de hemel

Contents

1 Introduction **9**

1.1 The concordance model of cosmology 10
1.2 Observational cosmology 10
1.3 Open questions 11
1.3.1 S_8 tension 13
1.3.2 $w_0 w_a$CDM 13
1.4 Fundamentals of cosmic shear 14
1.4.1 Deflection by a point mass 14
1.4.2 Lensing by an extended matter distribution 15
1.4.3 Approximations in weak lensing 17
1.4.4 Tomographic lensing 18
1.5 Weak lensing surveys 19
1.6 Weak lensing statistics 20
1.6.1 Shear two-point statistics 20
1.6.2 Higher-order weak lensing statistics 21
1.6.3 Weak lensing peaks 22
1.7 Cosmological simulations 23
1.7.1 N-body simulations 24
1.7.2 Hydrodynamical simulations 25
1.7.3 Subgrid physics 25
1.7.4 FLAMINGO 26
1.8 Baryonic impact on weak lensing 27
1.9 Matter power spectrum reconstruction 28
1.10 This thesis 29

References **31**

2 The FLAMINGO project: Baryonic impact on weak gravitational lensing convergence peak counts **35**

2.1 Introduction 36
2.2 Weak Lensing Theory 39
2.3 Methods 42
2.3.1 The FLAMINGO simulations 42
2.3.2 Model variations 44

2.3.3 Lightcones . . . 47
2.3.4 Full-sky WL convergence maps . . . 48
2.3.5 Angular power spectra . . . 52
2.4 Results . . . 54
2.4.1 Numerical convergence . . . 55
2.4.2 Cosmic variance . . . 56
2.4.3 Comparison of fiducial dark matter only and hydrodynamical models . . . 58
2.4.4 Baryonic variations . . . 61
2.4.5 Cosmology variations . . . 65
2.4.6 Separability of cosmological and hydrodynamical effects . 68
2.5 Discussion . . . 70
2.6 Summary and Conclusions . . . 72

References **75**

Appendix **78**
2.A Box rotation . . . 78
2.B Smoothing and Noise . . . 79

3 The FLAMINGO project: cosmology with the redshift dependence of weak gravitational lensing peaks **83**
3.1 Introduction . . . 84
3.2 Theory . . . 87
3.3 Methods . . . 88
3.3.1 FLAMINGO . . . 88
3.3.2 Weak-lensing peaks . . . 92
3.3.3 Halo light-cone . . . 98
3.3.4 Halo–peak matching . . . 98
3.3.5 Observed redshifts . . . 101
3.4 Results . . . 103
3.4.1 Peak distribution in halo mass . . . 103
3.4.2 Completeness . . . 104
3.4.3 Redshift distribution of SNR > 5 WL peaks . . . 108
3.4.4 Baryonic and cosmology impact . . . 110
3.4.5 Cosmology with the redshift of WL peaks . . . 117
3.5 Summary and Conclusions . . . 118

References **121**

Appendix **124**
3.A Smoothing . . . 124
3.B Noise peak distribution . . . 125
3.C Halo–peak matching characteristics . . . 127

4 Cosmological constraining power of the redshifts, heights, and angular clustering of weak gravitational lensing peaks 129
4.1 Introduction . . . 130
4.2 Theory . . . 132
4.3 Simulations . . . 134
4.4 Methods . . . 136
4.4.1 Smoothing . . . 137
4.4.2 Peak statistics . . . 137
4.4.3 Two-point correlation functions . . . 139
4.4.4 Inference forecast . . . 139
4.5 Results and Discussion . . . 141
4.5.1 The impact of cosmology . . . 143
4.5.2 Cosmological constraints forecast . . . 144
4.5.3 Complementarity of WL peak statistics . . . 151
4.5.4 Comparison to shear two-point correlation functions . . . 152
4.5.5 Impact of the smoothing scale . . . 154
4.6 Conclusions . . . 160

References 163

Appendix 166
4.A Emulator construction and validation . . . 166
4.A.1 Redshift distribution . . . 166
4.A.2 Height distribution . . . 167
4.A.3 Angular clustering . . . 167
4.A.4 Leave-one-out test . . . 167
4.B Correlation matrix . . . 169
4.C Cosmological constraints from WL peaks . . . 171
4.D Impact of individual parameters . . . 172

5 First step toward matter power spectrum reconstruction with Stage III weak gravitational lensing surveys 179
5.1 Introduction . . . 180
5.2 Theory . . . 183
5.3 Data . . . 184
5.4 Methodology . . . 187
5.5 Results . . . 187
5.5.1 Double power law constraints . . . 187
5.5.2 Comparison to ΛCDM predictions . . . 192
5.5.3 Suppression of the matter power spectrum . . . 192
5.6 Conclusions . . . 193

References 196

Appendix 198

5.A KiDS-1000 COSEBIs prediction 198

6 Matter power spectrum reconstruction with KiDS-Legacy 201
6.1 Introduction 202
6.2 Data and analysis 203
6.3 Results and discussion 204
6.3.1 Matter power spectrum constraints 204
6.3.2 Internal consistency with ΛCDM 206
6.3.3 Small-scale suppression 207
6.4 Conclusions 209
6.5 Data availability 210

References 211

Appendix 212
6.A Acknowledgements 212
6.B Method details 213
6.B.1 Regularised deprojection 213
6.B.2 Double power law 215
6.C Deprojection correlation matrix 215
6.D Posterior predictive distribution 216

Nederlandse samenvatting 219
Introductie 219
Dit proefschrift 220

Publications 223

Curriculum Vitae 227

Acknowledgements 229

1 Introduction

The most fundamental questions mankind has asked itself are about our origin and place in the Universe. Where do we come from? What is the Universe made of? How did it evolve? These are the questions at the heart of cosmology and may at first glance seem more philosophical in nature, but in reality, these questions can already be answered to an incredible amount of detail. Our current best theory can accurately describe the Universe from just a fraction of a second after the Big Bang all the way to today, some 13.8 billion years later. In the meantime, the theory can describe accurately many physical processes such as Big Bang nucleosynthesis, the formation of light elements in the first minutes after the Big Bang, recombination, the time when the Universe cooled enough to allow electrons and hydrogen to recombine and create a neutral Universe, reionisation, the ionisation of cosmic hydrogen from light produced by the first luminous objects, and structure formation, including the formation of stars, galaxies, clusters, and the cosmic web.

The field of cosmology was revolutionised by Einstein's formulation of the theory of general relativity (GR; Einstein, 1916, 1917). Combining Einstein's theory with the cosmological principle, meaning that at large enough scales the Universe is homogeneous – the same everywhere – and isotropic – the same in all directions – the Universe can be described in terms of the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, as (Friedmann, 1922, 1924; Lemaître, 1931; Robertson, 1935, 1936; Walker, 1937)

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + R^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right), \tag{1.1}$$

where $\mathrm{d}s^2$ is the infinitesimal line element that describes distances in space, c is the speed of light in vacuum, t is cosmic time, R is the radius of the Universe, r is the radial coordinate as seen from an observer, $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ are the angular coordinates, as seen from the same observer, and k describes the curvature of the Universe. It is conventional to normalise the radial coordinates by the current size of the Universe, R_0, which defines the scale factor, a, as $a(t) = R(t)/R_0$, which encodes the past expansion of the Universe. Eq. 1.1, which emerges from only assuming GR and the cosmological principle, is the backbone of modern

1

cosmology; it allows us to describe spacetime on cosmological scales, model the growth of the Universe, and connect its energy content to the expansion history.

1.1 The concordance model of cosmology

Since the formulation and development of the theory of GR, a single model has gradually established itself as the standard model of cosmology. This model, referred to as ΛCDM, describes a universe governed by three primary components. The dominant component in terms of energy density is dark energy (DE, Λ if its energy density is constant with cosmic time). DE is a so-far-unknown quantity that is needed to describe the accelerated expansion as first demonstrated by type Ia supernovae (Riess et al., 1998; Perlmutter et al., 1999) and through cosmic microwave background (CMB) anisotropies (Spergel et al., 2003). Preceding these measurements, galaxy clustering and cluster abundance measurements had already shown that the matter energy density parameter, $\Omega_{\rm m}$, which is the ratio of the matter energy density to the critical energy density of the Universe, $\rho_{\rm m}/\rho_{\rm crit}$, where $\rho_{\rm crit}$ is the energy density required for a spatially flat Universe, equals $\Omega_{\rm m} \approx 0.3$. The cluster and clustering measurements had thus shown that there is insufficient matter to accommodate a spatially flat universe (e.g., Maddox et al., 1990; Efstathiou et al., 1992; White et al., 1993). Later, DE was also shown to be required to directly explain the observations of the integrated Sachs-Wolfe effect (Boughn & Crittenden, 2004) and baryonic acoustic oscillations (BAO; Eisenstein et al., 2005).

The second major component is cold dark matter (CDM). The first evidence for CDM already dates back to the 1930s, when Zwicky (1933) found that the visible mass of galaxy clusters was insufficient to explain their dynamics. In this context, 'cold' refers to the thermal velocities of the particles at decoupling, which sets the free-streaming length and therefore impacts structure formation. The preference for CDM was established as it more accurately reproduced clustering signals than hotter DM models in cosmological simulations (Peebles, 1982; Blumenthal et al., 1984). As for DE, we have no solid understanding of the physical nature of DM.

The final component, baryons, contains all electromagnetically interacting matter we know of, including atoms, molecules, dust, comets, planets, stars, and gas clouds. Combined, these only make up ~ 5 per cent of the Universe's energy density.

1.2 Observational cosmology

Observational cosmology is the branch of astronomy that analyses the content, structure, and evolution of the Universe through observations. Some key cosmological observables, as a function of the spatial scales and cosmic times they

probe, are illustrated in Fig. 1.1. These include the clustering of galaxies (gg; e.g., Reid et al., 2010) and quasi-stellar objects (QSOs; e.g., DESI Collaboration et al., 2025b), which quantifies the average excess probability of objects in the vicinity of each other compared to a random distribution, the CMB (e.g., Planck Collaboration et al., 2020b), the remnant radiation of the early hot Universe that has been free streaming through the Universe since recombination and reflects the conditions in the Universe just $\sim 380\,000$ years after the Big Bang, CMB lensing ($\Phi\Phi$; e.g., Lewis & Challinor, 2006), cosmic shear ($\gamma\gamma$; e.g., Hoekstra & Jain, 2008), the weak distortions of, respectively, the CMB or background galaxy images due to weak gravitational lensing (WL) by the large-scale structure (LSS) of the Universe, the Lyman-α forest (Lyα; e.g., Palanque-Delabrouille et al., 2015), a series of absorption features seen in the spectra of distant sources caused by the distribution of neutral hydrogen clouds in the Universe, and cluster abundance (CA; e.g., Tegmark et al., 2004), the number of most massive gravitationally bound structures in the Universe. Combined, these observables probe the distribution of matter across a range of different spatial scales from just a few hundred thousand years after the Big Bang to the present day.

1.3 Open questions

Despite being able to accurately predict all observables as mentioned in the previous section, the physical nature of the two main components of the ΛCDM model, DM, which accounts for ~ 25 per cent of the energy density, and DE, which accounts for ~ 70 per cent, remains elusive, and very little substantial progress has been made in the last few decades. Additionally, the very early stages of the Universe, inflation, are very poorly understood.

An attempt to predict the energy density of DE can be obtained from quantum field theory (QFT). In QFT, the energy density of the vacuum scales as $\rho_{\rm vacuum} \approx \Lambda^4/(16\pi^2)$, where Λ is the wave number cut-off for summing the zero-point normal modes of the vacuum field. If GR holds all the way to the Planck scale, $\Lambda \approx \sqrt{1/(8\pi G)}$, with G the gravitational constant, such that $\rho_{\rm vacuum} \approx 10^{110}$ erg cm^{-3} (e.g., Weinberg, 1989). In reality, the observed value is $\rho_\Lambda \approx 10^{-8}$ erg cm^{-3}. This ~ 120-order-of-magnitude wrong estimate is referred to as the cosmological constant problem or vacuum catastrophe, and it underscores the current incompatibility between QFT and GR.

Although we have no clear understanding of the origins of DM and DE, or of the physical processes underlying inflation, assuming their existence still allows us to predict a plethora of observational cosmological probes with great accuracy. However, when results from cosmological analyses of different observables are compared, slight differences emerge. These differences are referred to as 'tensions' and are a different kind of issue from not knowing the physical origin of the main model ingredients, as they indicate conflicting results rather than just

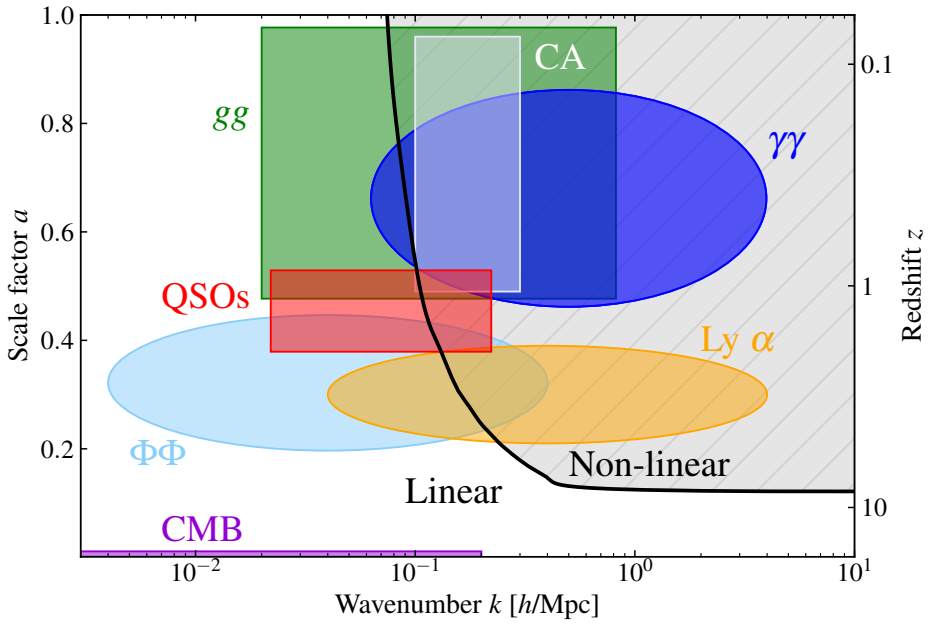


Figure 1.1: Different observational cosmological probes, along with the scales and times to which they are approximately sensitive. The solid black curve separates the non-linear from the linear collapse regime. From early to late time, the different probes are the CMB (purple), CMB lensing ($\Phi\Phi$, light blue), Lyman-α forest (Ly α, orange), quasi-stellar objects (QSOs, red), cosmic shear (blue, $\gamma\gamma$), galaxy clustering (gg, green), and cluster abundance (CA, white). The rectangles correspond to probes of which the radiation is directly observed, whereas for the ellipses, the studied process happens somewhere along the propagation through the Universe. Figure taken from Broxterman & Kuijken (2024).

our ignorance. This means there is an issue with our modelling of at least one of the measurements being compared. This may be a consequence of physical laws not being properly described in the vanilla ΛCDM model, for example, if DM is decaying, or of mismodelling or underestimating the impact of systematics. The two most prominent tensions in cosmology of the last decade are the H_0 tension and the S_8 tension. The first refers to the value of the Hubble parameter, a measure of the Universe's expansion rate. Here, the value from the local Universe, as measured by supernovae calibrated using a distance ladder, differs from that from the early Universe as seen by *Planck* (e.g., Kamionkowski & Riess, 2023).

1.3.1 S_8 tension

The second tension concerns a measure of the clumpiness or homogeneity of matter in the Universe. Here, several local LSS probes are in slight tension with the *Planck* CMB measurements (for a recent review, see Pantos & Perivolaropoulos, 2026). Improvements in the precision of cosmic shear observations first hinted at the discrepancy (Hildebrandt et al., 2017). It is most frequently presented in terms of the parameter $S_8 = \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$, where σ_8 is the root-mean-square fluctuation of the matter density field in spheres with radius $8\,h^{-1}\,\mathrm{Mpc}$ extrapolated to present-day using linear theory, which is a parameter that is well constrained by WL observations.

By now, the tension has been reported between *Planck* CMB results and results from WL (e.g. Secco et al., 2022; Dark Energy Survey and Kilo-Degree Survey Collaboration et al., 2023), galaxy clustering (e.g. Tröster et al., 2020; Philcox & Ivanov, 2022), redshift space distortions (e.g. Kazantzidis & Perivolaropoulos, 2018; Benisty, 2021), cluster counts (e.g. Bocquet et al., 2019; Abbott et al., 2020) or cross-correlations between different probes (e.g. Heymans et al., 2021; García-García et al., 2021; White et al., 2022). However, not all local probes report tensions, as seen, for instance, in the eBOSS quasar power spectrum inference from Chudaykin & Ivanov (2023), or in some distance tracers, such as the Tully-Fisher relation, fundamental plane, and supernovae (e.g. Stiskalek, 2025). The most recent eROSITA cluster abundance inference even finds slightly larger S_8 values than *Planck* (Ghirardini et al., 2024).

Whereas CMB theory is very linear, as shown by the black curve in Fig. 1.1 that separates the linear and non-linear collapse regimes, and therefore relatively simple, the low-redshift observables probe smaller scales associated with non-linear gravitational collapse. This is a blessing and a curse, as the non-linear collapse regime includes a wealth of additional information, but at the cost of having to deal with many more modelling options and choices. This has motivated some to suggest that inaccurate modelling in the non-linear regime leads to the observed tensions (e.g., Amon & Efstathiou, 2022).

Although the tension gained traction due to the cosmic shear observations presented in Hildebrandt et al. (2017), the latest cosmological inference from the same survey no longer reports this tension (Stölzner et al., 2025; Wright et al., 2025). Therefore, whether there actually is a discrepancy, or whether systematics are misestimated or their uncertainties are underestimated, remains to be seen.

1.3.2 w_0w_aCDM

A recent extension of the ΛCDM model that has received much attention is the w_0w_aCDM model. This model accounts for the evolution of DE as parametrised

1

by the equation of state parameter

$$w_{\rm DE} = w_0 + w_a(1-a), \tag{1.2}$$

where $w_{\rm DE} \equiv P_{\rm DE}/\rho_{\rm DE}$ is the equation of state parameter for DE and w_0 and w_a follow the Chevallier-Polarski-Linder (CPL) parametrisation (Chevallier & Polarski, 2001; Linder, 2003). When $w_0 = -1$ and $w_a = 0$, the model reduces to ΛCDM. The parametrisation has gained a lot of traction as the measurements of the distribution of galaxies through their clustering and BAO signal as seen by the first couple of years of observations of the multi-fibre spectrograph Dark Energy Spectroscopic Instrument (DESI) seem to prefer values of w_0 and w_a inconsistent with ΛCDM (DESI Collaboration et al., 2025a,c).

Although the CPL parametrisation is a popular extension of ΛCDM, there is no theoretical basis for expecting DE to evolve in this manner. The parametrisation arises from a first-order Taylor expansion of the DE evolution at the present time. Since we already don't know what DE is, constraining its properties in this manner is a purely phenomenological approach that can, at best, offer hints about the underlying physical process driving any evolution. Nevertheless, we can expect to achieve much tighter constraints on these parameters from all Stage IV surveys within the next decade. For example, in the study definition report of ESA's *Euclid* mission, which aims to determine the nature of DE and DM, measuring the values of w_0 and w_a to within 1σ standard deviations of 0.02 and 0.1, respectively, is listed as its first main scientific objective (Laureijs et al., 2011).

1.4 Fundamentals of cosmic shear

As this thesis focuses on the analysis of cosmic shear signals in simulations and observations, this section provides a brief historical overview of WL and derives key equations that form its theoretical foundation. For extended reviews on both parts, see Bartelmann & Schneider (2001) and Kilbinger (2015). In what follows, the universe is assumed to have negligible curvature such that the comoving angular diameter distance, $f_k(\chi)$, equals the comoving line of sight distance, χ, as is found in the combination of *Planck* and BAO measurements to within 0.2% precision (Planck Collaboration et al., 2020a).

1.4.1 Deflection by a point mass

Alongside the prediction of the precession of the perihelion of Mercury and gravitational redshift, one of Einstein's initial proposed tests of GR was that the deflection of light by a massive body should be twice as large as the classical Newtonian value, which is obtained by assuming photons are massive particles (Einstein, 1911, 1916). The prediction can be derived by considering a weakly perturbed ($\phi \ll c^2$) Minkowski metric, which describes the evolution of a test parti-

cle around a static source in a universe where the stress-energy is invariant under spatial rotations (i.e., a universe without anisotropic stress). In the Newtonian gauge, the FLRW infinitesimal metric line element, $\mathrm{d}s^2$, as seen in Eq. 1.1, can be expressed in terms of the Newtonian gravitational potential, ϕ, as

$$\mathrm{d}s^2 = -\left(1+\frac{2\phi}{c^2}\right)c^2\,\mathrm{d}t^2 + a(t)^2\left(1-\frac{2\phi}{c^2}\right)\left[\mathrm{d}\chi^2+\chi^2\mathrm{d}\Omega\right]. \tag{1.3}$$

The reference frame may be chosen such that the photon is moving radially towards the observer, in which case its angular position remains unchanged ($\mathrm{d}\Omega^2 = 0$). By setting $a = 1$ as the present, such that the comoving distance, χ, equals the physical distance, r, using the fact that light travels on null geodesics ($\mathrm{d}s^2 = 0$), and the binomial approximation ($[1+x]^n \approx 1+nx$ for $x \ll 1$), the light travel time of a photon is given by an integral over the photon path

$$t = \frac{1}{c}\int_{\mathrm{path}} \mathrm{d}\chi\left(1-\frac{2\phi}{c^2}\right). \tag{1.4}$$

The deflection angle ($\hat{\boldsymbol{\alpha}}$) can be derived by considering Fermat's principle of minimal light travel time, i.e., $\delta t = 0$, to give a set of Euler-Lagrange equations (e.g. Schneider, 1985; Blandford & Narayan, 1986), which can be integrated along the photon path to yield the angle at which the photon is deflected,

$$\hat{\boldsymbol{\alpha}} = -\frac{2}{c^2}\int_{\mathrm{path}} \mathrm{d}\chi\,\nabla_\perp\phi\,, \tag{1.5}$$

where the gradient is evaluated perpendicular to the photon's direction of travel. This deflection is twice the prediction from Newtonian gravity (von Soldner, 1804). In 1919, the English astronomer Eddington led an effort to image the positions of stars near the Sun during a total solar eclipse, and the resulting observations were used to confirm Einstein's prediction (Dyson et al., 1920).

1.4.2 Lensing by an extended matter distribution

The first proposal to use entire galaxies as lenses to probe their masses was put forward by Zwicky (1937). To describe this phenomenon, the previous equations need to be extended to account for a general matter distribution rather than a point mass. The main quantities relevant to describing the WL formalism are depicted in Fig. 1.2.

The mechanism can be described by considering the paths of two neighbouring light rays that are separated by a transverse comoving separation $\boldsymbol{x}$, as seen by an observer. Given a gravitational lens at distance χ', the deflection angle induced by this lens is

$$\mathrm{d}\hat{\boldsymbol{\alpha}} = -\frac{2}{c^2}\int_{\mathrm{path}} \mathrm{d}\chi'\,\nabla_\perp\phi[\boldsymbol{x}(\chi'),\chi']. \tag{1.6}$$

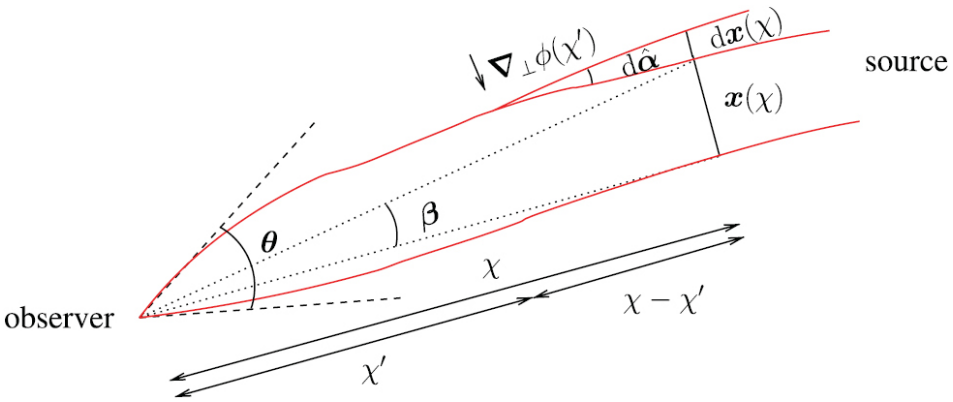


Figure 1.2: Main quantities relevant to describing cosmic shear. The paths of two light rays, separated by a transverse comoving distance $\mathbf{x}$, are deflected by the transverse gradient $\nabla_\perp\phi$ of the gravitational potential at the comoving distance χ'. The dashed lines, with an angular separation $\boldsymbol{\theta}$, illustrate the apparent directions of the light rays, as seen by the observer, whereas the dotted lines are the unperturbed photon paths, corresponding to the angle $\boldsymbol{\beta}$ that the unperturbed transverse separation $\boldsymbol{x}$ would subtend. Figure taken from Kilbinger et al. (2017).

The change in the transverse separation between two incoming light rays, as they experience a non-homogeneous matter distribution, as seen from the lens, is

$$\mathrm{d}\boldsymbol{x} = (\chi - \chi')\,\mathrm{d}\hat{\boldsymbol{\alpha}}. \tag{1.7}$$

The transverse comoving separation, which is achieved by integrating over lenses at all distances until the source at distance χ, is given by

$$\boldsymbol{x} = \chi\boldsymbol{\theta} - \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi'\,(\chi - \chi')\left\{\nabla_\perp\phi^{(1)}[\boldsymbol{x}(\chi'),\chi'] - \nabla\phi^{(2)}(\chi')\right\}, \tag{1.8}$$

where $\boldsymbol{\theta}$ is the observed angular separation, such that the first term represents the transverse distance at emission and the second term represents the change caused by gravitational lensing. The superscripts $^{(1)}$ and $^{(2)}$ indicate the two different light rays. As such, the observed angular separation can be related to the relative unlensed position, $\boldsymbol{\beta}$, via the total deflection angle $\boldsymbol{\alpha}$ as

$$\boldsymbol{\beta} = \frac{\boldsymbol{x}}{\chi} = \boldsymbol{\theta} - \boldsymbol{\alpha}, \tag{1.9}$$

which is commonly referred to as the lens equation. The deflection angle is then given by

$$\boldsymbol{\alpha} = \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi'\,\frac{\chi - \chi'}{\chi}\nabla_\perp\left\{\phi^{(1)}[\boldsymbol{x}(\chi'),\chi'] - \phi^{(2)}(\chi')\right\}. \tag{1.10}$$

Alternatively, the equations are expressed in terms of the lensing potential, ψ, as

$$\psi(\boldsymbol{\theta}, \chi) = \frac{2}{c^2} \int_0^{\chi} \mathrm{d}\chi' \frac{\chi - \chi'}{\chi \chi'} \phi[\boldsymbol{x}(\chi'), \chi']. \tag{1.11}$$

The derivative with respect to comoving coordinates is related in terms of angular derivatives as $\nabla_\perp = \nabla_{\boldsymbol{\theta}}/\chi$, such that the lensing equation reduces to

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \nabla_{\boldsymbol{\theta}} \psi(\boldsymbol{\theta}, \chi), \tag{1.12}$$

where the term relating to ray $^{(2)}$ in Eq. 1.10 has vanished as it does not depend on $\boldsymbol{\theta}$ and the superscript $^{(1)}$ has been dropped for the first term.

1.4.3 Approximations in weak lensing

As deflections remain small, results are typically presented by linearising the lens equation, i.e., expanding it to first order in the potential. The results can be presented as the derivative of the deflection angle in the lensing matrix/lensing Jacobian/magnification matrix, A as

$$\frac{\partial \beta_i}{\partial \theta_j} \equiv A_{ij} = \delta_{ij} - \partial_i \partial_j \psi, \tag{1.13}$$

where i and j represent the two angular derivatives perpendicular to the photon path. The magnification matrix can be decomposed into a scalar component that quantifies the isotropic magnification, κ, referred to as the WL convergence, and a spin-two field that measures the anisotropic stretching, γ, the WL shear as

$$A = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \tag{1.14}$$

Another approximation that is commonly made is to evaluate the separation distance in the previous expressions to the zeroth order, $\boldsymbol{x}(\chi') = \chi' \boldsymbol{\theta}$, i.e., evaluating the potential at the unperturbed photon path. This approximation, known as the Born approximation, is applied to the subsequent equations. By combining equations 1.11 – 1.14 the WL convergence is equivalent to

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{1}{c^2} \int_0^{\chi} \mathrm{d}\chi' \frac{\chi - \chi'}{\chi' \chi} \nabla^2_{\boldsymbol{\theta}} \phi(\chi' \boldsymbol{\theta}, \chi'). \tag{1.15}$$

In ΛCDM, we can use the 3-dimensional Poisson equation, $\nabla^2 \phi = 4\pi \overline{\rho_\mathrm{m}} a^2 G \delta$, to relate the Newtonian potential to the matter field. δ is the matter density contrast, given by

$$\delta = \frac{\rho_\mathrm{m} - \overline{\rho_\mathrm{m}}}{\overline{\rho_\mathrm{m}}}, \tag{1.16}$$

1

where $\overline{\rho_\mathrm{m}}$ is the mean matter density of the universe. We assume that varying contributions along the line of sight cancel such that the 3-dimensional Laplacian reduces to $\nabla^2 = \nabla_\perp^2 + \partial^2/\partial\chi^2 \approx \nabla_\perp^2 = \nabla_{\boldsymbol{\theta}}^2/\chi^2$, and

$$\nabla_{\boldsymbol{\theta}}^2\phi = \chi^2 \frac{3H_0^2\Omega_\mathrm{m}}{2a}\delta, \tag{1.17}$$

which shows that WL measures the deflection due to all matter in the Universe. This illustrates a key advantage of WL. Whereas many observational cosmological probes, such as the thermal and kinetic Sunyaev-Zel'dovich (SZ) effect, the dispersion measure of fast radio bursts, or X-ray observations, only provide information about the distribution of baryons, WL probes the total matter distribution. Substituting Eq. 1.17 in 1.15 allows to estimate the WL convergence as

$$\kappa(\boldsymbol{\theta},\chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2}\int_0^\chi \mathrm{d}\chi' \frac{(\chi-\chi')\chi'}{\chi a(\chi')}\delta(\chi'\boldsymbol{\theta},\chi'). \tag{1.18}$$

This equation also illustrates how to estimate WL signals in cosmological simulations, as these simulations track the evolution of the matter field, which can be related to the density contrast field via Eq. 1.16.

1.4.4 Tomographic lensing

By now, the formalism has been extended from the deflection of a single source by a single lens to that by an arbitrary mass distribution. When an ensemble of source galaxies is observed, the measured signal can be described as an integral over the normalised source distribution, $n_\mathrm{s}(\chi)$, as

$$\kappa(\boldsymbol{\theta}) = \int_0^{\chi_\mathrm{hor}} \mathrm{d}\chi\, n_\mathrm{s}(\chi)\,\kappa(\boldsymbol{\theta},\chi), \tag{1.19}$$

where χ_hor is the comoving distance until the outer edge of the survey. To gain more information about the 3-dimensional distribution of matter, sources can be split into samples with different redshift distributions and combined in a single analysis, a technique also referred to as tomographic lensing. By taking the square of Eq. 1.18 in Fourier space and applying the (extended) Limber approximation (Limber, 1953; LoVerde & Afshordi, 2008), the WL convergence angular power spectrum as a function of multipole moment, $\mathcal{C}(\ell)$, can be related to the power spectrum of the density contrast, i.e. the matter power spectrum, $P_\mathrm{m}(k) = \left\langle|\hat{\delta}(k)|^2\right\rangle$, with $\hat{\delta}$ being the Fourier transform of the matter density contrast, as

$$\mathcal{C}(\ell) = \frac{9\Omega_\mathrm{m}^2 H_0^4}{4c^4}\int_0^{\chi_\mathrm{hor}} \mathrm{d}\chi \frac{W^2(\chi)}{a^2(\chi)} P_\mathrm{m}\left[\frac{\ell+1/2}{\chi}, z(\chi)\right], \tag{1.20}$$

with the window function

$$W(\chi) = \int_\chi^{\chi_\mathrm{hor}} \mathrm{d}\chi' n_\mathrm{s}(\chi')\frac{\chi'-\chi}{\chi'}. \tag{1.21}$$

1.5 Weak lensing surveys

The first detections of cosmic shear were independently made by 4 groups only $\approx$ 25 years ago, each measuring tens of thousands of galaxies over areas of up to a few square degrees on the sky (Bacon et al., 2000; Kaiser et al., 2000; Van Waerbeke et al., 2000; Wittman et al., 2000).

Subsequent lensing surveys are currently typically grouped and referred to as different 'Stage' surveys. This nomenclature originates from a white paper by the Dark Energy Task Force (DETF; Albrecht et al., 2006).

The first stage, defined as completed at the time of writing the white paper, is characterised by small areas and shallow depths. These surveys demonstrated the feasibility of exploiting cosmic shear in a cosmological context, thereby serving as a proof of concept for WL. These years were characterised by developments of the shape measurement technique, crucial for obtaining unbiased estimates of cosmic shear (e.g., Heymans et al., 2006). Examples of surveys belonging to this group are COMBO-17 (Wolf et al., 2003), the Deep Lens Survey (Wittman et al., 2002), or from observations obtained at the Canada-France-Hawaii Telescope (CFHT; Van Waerbeke et al., 2001).

The second stage, consisting of ongoing surveys at the time of writing of the white paper, already used millions of galaxies, providing the first competitive cosmological constraints while also improving methods such as photo-z estimation, the procedure to estimate galaxy redshifts from flux measurements in a limited number of photometric bands, as estimating the redshift of each galaxy individually becomes less effective than using statistical clustering methods that describe the entire population (e.g., Lima et al., 2008). Examples of Stage II WL surveys are COSMOS (Leauthaud et al., 2007), the CFHT Lens Surveys (Heymans et al., 2012), and the Red-sequence Cluster Survey-2 (Gilbank et al., 2011).

The third generation of lensing surveys comprised future projects that had already been proposed and which were intended to be near-term and at medium cost. The current most notable Stage III surveys that have just finalised or are expected to finish their legacy analyses in the near future are the Kilo-Degree Survey (KiDS; de Jong et al., 2013), the Dark Energy Survey (DES; The Dark Energy Survey Collaboration, 2005), and the Hyper Suprime-Cam (HSC; Miyazaki et al., 2012). This generation is characterised by expanding up to thousands of square degrees, needing collaborations of tens of people to analyse all data, while also reaching the precision that requires the modelling of new systematics such as baryonic feedback (e.g., Semboloni et al., 2011) and intrinsic alignment of galaxies (e.g., Hirata et al., 2007).

The focus currently shifts from Stage III to Stage IV surveys. Although initially defined to encompass other surveys, WL analyses currently typically refer to three primary Stage IV WL surveys. Their analyses are underway, as *Euclid* is carrying out its first cosmological inference (Euclid Collaboration: Mellier et al.,

1

2025), *Rubin*'s LSST has started (LSST Science Collaboration et al., 2009), and *Roman* has launched (Spergel et al., 2015). Combined, these surveys will exhaust nearly all of the sky best suited to WL ($\sim 15\,000$ square degrees), aiming to provide sub-per cent-level constraints on the equation-of-state parameters of DE (w_0, w_a), the neutrino mass ($\sum m_\nu$), and modified gravity, while being systematics-limited.

Although not defined in the DETF white paper, several possibilities exist for Stage V lensing surveys. However, these will require new investments and facilities and are therefore not expected to be operational within the next decade. Examples of Stage V surveys are large ($\sim$ 30m) ground-based telescopes with wide fields to allow precise and deep measurements over large fields, spectroscopic WL, i.e., combining WL with spectroscopic surveys to get a spectroscopic redshift for each galaxy, or further increasing the wavelength coverage, for example, down to radio frequencies (Maartens et al., 2015).

1.6 Weak lensing statistics

1.6.1 Shear two-point statistics

In their principal analyses, Stage III collaborations typically summarise their data using two-point statistics, which can be estimated from the ellipticities of observed galaxies. The most commonly used two-point statistic is the shear two-point correlation function (2PCF), $\xi(\theta)$, which quantifies the relation between shear values at different positions as a function of their angular separation, θ. It is obtained by multiplying the observed ellipticities of galaxy pairs and averaging over a large sample. Given that shear is a spin-2 field, three independent 2PCF combinations exist. One of these vanishes in a parity-symmetric universe. The remaining two are typically presented as ξ_+ and ξ_- as

$$\begin{aligned}\xi_+(\theta) &= \langle\gamma_t\gamma_t\rangle(\theta) + \langle\gamma_\times\gamma_\times\rangle(\theta);\\ \xi_-(\theta) &= \langle\gamma_t\gamma_t\rangle(\theta) - \langle\gamma_\times\gamma_\times\rangle(\theta),\end{aligned} \tag{1.22}$$

where γ_t and $\gamma_\times$ are, respectively, the tangential component and cross component of the observed shear, which are rotated by 90 degrees. Theoretically, the shear-2PCF can be estimated from the angular power spectrum, as given in Eq. 1.20, by assuming a flat sky and vanishing B-modes as (Kilbinger et al., 2017)

$$\begin{aligned}\xi_+(\theta) &= \int_0^\infty \frac{\mathrm{d}\ell\,\ell}{2\pi}\mathcal{J}_0(\ell\theta)\mathcal{C}(\ell);\\ \xi_-(\theta) &= \int_0^\infty \frac{\mathrm{d}\ell\,\ell}{2\pi}\mathcal{J}_4(\ell\theta)\mathcal{C}(\ell),\end{aligned} \tag{1.23}$$

where $\mathcal{J}_0$ and $\mathcal{J}_4$ are cylindrical Bessel functions of the first kind of the zeroth and fourth order, respectively.

Other two-point statistics exist, such as band power spectra (e.g., Schneider et al., 2002) or the complete orthogonal set of e/b integrals (e.g., Schneider et al., 2010), that probe the same information. Some analyses opt for these statistics because they are computationally less expensive, sensitive only to a specific set of scales, or exhibit weaker correlations between data points. However, theoretically, like the shear-2PCFs, they remain projections of the angular power spectra, just with different window functions. An advantage of shear-2PCFs is that they are straightforwardly interpretable, as the quantity is measured as a function of the physical quantity angular separation. The others are measured as functions of unobservable parameters that may not directly correspond to a specific scale but rather to an average over different scales, and therefore, they miss some of the basic intuitive interpretation of shear-2PCFs.

1.6.2 Higher-order weak lensing statistics

As Stage IV WL surveys will measure the cosmic shear signal with great precision, a wealth of cosmological information will be contained in their measured signals. The commonly used two-point statistics, which, by definition, capture only the underlying Gaussian signal, do not capture all information about the non-linear growth of structure formation.

To extract the beyond-Gaussian cosmological information from WL signals, a suite of statistics, collectively referred to as higher-order weak lensing statistics (HOWLS), has been proposed. A non-exhaustive list includes three- and four-point correlation functions (e.g., Takada & Jain, 2003; Silvestre-Rosello et al., 2025), the one-point probability distribution (e.g., Taruya et al., 2002; Barthelemy et al., 2021), higher-order moments (e.g., the kurtosis and skewness; Van Waerbeke et al., 2013), peaks (e.g., Marian et al., 2009), voids (e.g., Betancort-Rijo et al., 2009), Minkowski functionals (e.g., Kratochvil et al., 2012), Betti numbers (Feldbrugge et al., 2019), scattering transform coefficients (e.g. Cheng et al., 2020), wavelet phase harmonics (e.g., Allys et al., 2020), or persistent homology (e.g., Heydenreich et al., 2021).

By combining several of these statistics, the tightest cosmological constraints can be obtained, while also breaking degeneracies between parameters (e.g., Euclid Collaboration: Ajani et al., 2023). When comparing several of these statistics, nearly all the information (and more) contained in the standard two-point inferences is also captured by HOWLS (Euclid Collaboration: Vinciguerra et al., 2025).

In principle, even more information can be extracted by not condensing the observed field into summary statistics but by instead carrying out a field-level inference (e.g., Boruah et al., 2022; Porqueres et al., 2022). Whereas these approaches could extract all, and more, of the information contained in HOWLS, it has yet to be demonstrated that they remain robust in the presence of non-

1

homogeneous survey properties, survey masks, noise correlations, uncertainties in point-spread function (PSF) modelling, high-dimensional and non-Gaussian likelihoods, complex non-linear and non-local transformations, for example the conversion from observed shear to the convergence that is usually estimated in simulations, and other unknown systematics.

HOWLS aims to serve as a bridge between classical two-point analyses and full non-linear field-level inferences. On the one hand, it furthers the goal of achieving the tightest possible (cosmological) constraints, while on the other hand, it provides statistics that can still be interpreted, for example, by measuring signals across a limited set of scales to isolate physical processes. Nevertheless, some interpretability is still lost when combining multiple statistics or, for example, when considering abstract topological statistics, especially when machine learning approaches are employed (e.g., Lucas Makinen et al., 2025).

1.6.3 Weak lensing peaks

As Chapters 2 – 4 focus on the properties of a single non-Gaussian statistic, this section zooms in on this statistic, WL peaks. WL peaks are the local maxima in WL convergence fields. The highest peaks correspond to the most massive structures in the Universe (Hamana et al., 2004). Whereas WL peaks are a non-Gaussian statistic that probes the non-linear gravity regime, as they are typically identified on WL convergence maps, their measurement is only first-order in the observed shear, and the statistic therefore potentially suffers less from systematics than other HOWLS. This is exemplified by the analysis of Gatti et al. (2024), who study the impact of source clustering, the correlation between galaxy number density and lensing signal as a consequence of galaxies tracing the underlying matter field, on several HOWLS and find that for most, the impact is significant, whereas for WL peaks, it is expected to have only a minor impact.

The cosmological potential of WL peaks has been studied in terms of the number of peaks and height distribution (e.g., Jain & Van Waerbeke, 2000; Wang et al., 2009; Kratochvil et al., 2010; Dietrich & Hartlap, 2010), their clustering (e.g., Marian et al., 2013; Davies et al., 2022), steepness (Li et al., 2023), and redshift distribution (Broxterman et al., 2025). Together, these statistics aim to probe the number, shape, and spatial distribution of (massive) haloes (e.g., Yang et al., 2011; Liu & Haiman, 2016).

For many of the non-Gaussian statistics, there is no satisfactory theoretical framework to predict the observable as a function of cosmology. For example, for WL peaks, the lowest-valued peaks can be predicted by describing the LSS as a Gaussian random field (Bardeen et al., 1986). Whereas this approach can reasonably well describe the value of negative peaks (Maturi et al., 2010), i.e., the local maxima in generally underdense regions, it fails to capture non-linear structure formation and therefore breaks down at higher WL convergence values. Anal-

ogously, the highest-valued peaks can be predicted by assuming all peaks are generated by a separate single massive structure along the line-of-sight, which allows for predicting the number of peaks using the halo mass function (HMF), i.e., the number of haloes of given mass existing in a certain volume of the Universe, and halo shape (Hamana et al., 2004). Whereas the assumption that these highest peaks are caused by single objects holds, the approach will fail for lower-valued peaks where projection effects or fluctuations due to the random shape noise start to dominate (Broxterman et al., 2025). The accuracy of many of these theoretical approaches is therefore not guaranteed, especially given the expected per cent-level precision goal of Stage IV WL surveys. For example, for the analytical prediction of WL peaks, as discussed in the previous paragraph, halo finders, which find and characterise structures in cosmological simulations, as well as the accuracy of emulators, already impact the HMF at the beyond per cent level (Kugel et al., 2025; Forouhar Moreno et al., 2025). Also, haloes are typically assumed to have a spherically symmetric profile (e.g., Navarro et al., 1997), while ΛCDM predicts haloes to be triaxial, which is also seen in cosmological hydrodynamical simulations (e.g., Velliscig et al., 2015; Chua et al., 2022). If the true shapes have a high degree of asphericity, the change in peak height and profile will depend on the alignment of the halo with the line of sight. Similarly, low-valued peaks can be plagued by projection effects and are highly sensitive to survey properties, particularly the shape noise (Broxterman et al., 2024), which must be well understood to achieve accurate constraints.

As no satisfactory theoretical predictions are generally available for these statistics, they need to be inferred using simulation-based inference (SBI). Here, SBI refers to a type of analysis where a suite of simulations is run with varying (cosmological) parameters. For each simulation, a prediction of a statistic is obtained by forward modelling a specific survey's measurements. By comparing observed statistics with simulations, often via emulators, probabilistic (cosmological) constraints can be obtained. Therefore, it is necessary to run a large number of cosmological simulations to calibrate the dependence of HOWLS statistics on cosmology. An additional advantage of this approach is that many systematics, e.g., survey geometry, varying depth, intrinsic alignment, and source clustering, can be more straightforwardly incorporated through forward modelling, although the current commonly used methods are not accurate enough to control all systematics at the per cent level (e.g., Lee et al., 2026).

1.7 Cosmological simulations

To assess the cosmological potential of different observables, such as the WL peak properties discussed in the previous section, astronomers rely on cosmological simulations. Astronomy, particularly in the fields of cosmology and galaxy formation, is a unique physical science in terms of its application of the scientific

1

method. Whereas a chemist may conduct experiments under controlled conditions in their laboratory, it is impossible for cosmologists to go out into the Universe and change its composition, evolution, or initial conditions. It is therefore impossible for cosmologists to follow the scientific method in its traditional form. The only thing we can do is try to understand the physical processes governing the Universe from images, spectra, and energy measurements taken from within the solar system.

A major tool that can serve as a laboratory for cosmologists is cosmological simulations. Instead of directly altering the composition of the Universe, we can program physical laws into large computer systems and have them predict the Universe's evolution and properties. In these mock universes, or cosmological simulations, we can then change the composition and the laws governing their evolution (e.g., heavier neutrinos, stronger magnetic fields, or the inclusion of exotic DM species). By comparing the outputs of these simulations, i.e., their predictions of the structure and evolution of the simulated universe, to actual observations, we can still study the content and evolution of the Universe.

Additionally, simulations can help by quantifying the impact of uncertain physical processes. In simulations, the strength of these processes can be systematically varied to study their impact on a set of observables. They may help quantify the physical scales that are (not) affected, provide predictions of the uncertainty in the impact, which can subsequently be marginalised over, and help identify statistics that are insensitive to these uncertain physical processes.

Finally, simulations provide an excellent platform for testing models. The simulations can be used to test and validate models, serving as a ground truth. Although they do not one-to-one correspond to the actual Universe, since they are designed to capture the effects of the dominant physical processes governing it, models can be tested to determine whether they uncover the ground truth of the simulations. For example, if a method cannot accurately recover the truth of this highly simplified universe, how can we trust that it will perform well when applied to real data?

1.7.1 N-body simulations

The simplest form of cosmological simulation only tracks the effect of gravity, the most important force on the largest scales, as it is a long-range force scaling as $\propto r^{-2}$, and there exists no such thing as negative masses, such that the gravitational force always adds up. In contrast, the electromagnetic force, which scales the same, vanishes on large scales as positive and negative charges cancel each other out. Simulations that only track gravity are referred to as N-body or dark-matter-only (DMO) simulations. The earliest simulations consisted of only tens to hundreds of particles (e.g., von Hoerner, 1960; Aarseth, 1963). These simula-

tions described very simple systems to study stellar dynamics or the evolution of galaxies in clusters.

Gradually, these simulations became more advanced by, for example, improving the integration order (e.g., Aarseth, 1966), introducing neighbouring schemes where the contribution of neighbouring particles is updated more often than that of distant particles (Ahmad & Cohen, 1973), and by developing methods for the construction of initial conditions necessary for cosmological simulations (Efstathiou et al., 1985). For cosmological analyses, large-volume simulations are necessary to probe a representative part of the Universe. By now, the simulations with the largest number of resolution elements trace the evolutions of over a trillion (10^{12}) components in Gpc^3 volumes, as seen in the Uchuu (Ishiyama et al., 2021), MillenniumTNG (Hernández-Aguayo et al., 2023), FLAMINGO-10K (Pizzati et al., 2024), *Euclid* Flagship 2 (Euclid Collaboration: Castander et al., 2025), and HyperMillennium (Wang et al., 2025) simulations.

1.7.2 Hydrodynamical simulations

DMO simulations are useful for studying the LSS on large scales, where gravity is the only dominant physical force. When smaller scales are studied, especially for the precision required for Stage IV surveys, additional physics must be accounted for. In these cases, baryonic physics cannot be ignored. Whereas DM can be modelled as a collisionless, pressureless fluid on large scales, gas particles can radiate energy and cool, and are typically modelled as a collisional, non-conducting, ideal gas. For recent reviews of cosmological hydrodynamical simulations of galaxy formation, see Vogelsberger et al. (2020) and Crain & van de Voort (2023).

1.7.3 Subgrid physics

When gas cools down enough, it should form stars. Because the resolution of cosmological simulations is far from sufficient to resolve the conditions for the formation of individual stars, assumptions must be made to approximate the underlying physical processes. For star formation, this can be achieved by stochastically converting gas particles into collisionless stellar particles, provided they are in sufficiently dense and/or cold enough environments (e.g., Springel & Hernquist, 2003). Such a model, which aims to represent a physical process occurring at scales smaller than are resolved in a simulation, is called a subgrid model. Other subgrid models that are relevant to cosmological hydrodynamical simulations are models for the cooling and heating of gas (e.g., Katz et al., 1996; Wiersma et al., 2009), supernova / stellar feedback (e.g., Katz, 1992; Kay et al., 2003), and black hole accretion and feedback from active galactic nuclei (e.g., Springel et al., 2005; Booth & Schaye, 2009). These subgrid prescriptions typically contain several free parameters, whose values may be chosen by the simula-

1

tor. Some of these values can be based upon theoretical arguments, for example, the star formation threshold density (Schaye, 2004). For other parameters, this cannot be done in a meaningful way. In those cases, simulators typically determine the values of these subgrid parameters through calibration, which can be explicit or implicit. The values of the subgrid parameters are chosen such that the simulations reproduce a set of key observations typically including a combination of the present-day galaxy stellar mass function, (hot) gas or baryon fractions of clusters, the stellar-to-halo mass relation, the galaxy size-mass relation, or the black hole-stellar mass relation (e.g., McCarthy et al., 2017; Henden et al., 2018; Springel et al., 2018; Davé et al., 2019; Dolag et al., 2025). Together, depending on the choice of statistics, calibrating to these observables ensures that the simulations yield sensible results for galaxy properties, gas/baryon content, or LSS.

1.7.4 FLAMINGO

As three chapters of this thesis focus on analysing outputs from cosmological simulations using a single model, the philosophy and details of these simulations are described in greater detail. These simulations are the FLAMINGO simulations (Schaye et al., 2023; Kugel et al., 2023). The flagship FLAMINGO simulation is among the hydrodynamical simulations with the largest number of resolution elements, namely $2 \times 5040^3 + 2800^3$.

As explained in the previous section, cosmological hydrodynamical simulations are calibrated to a set of observables. FLAMINGO is unique in that it is the first suite of simulations to use machine learning, specifically Gaussian process emulators, to calibrate to the $z = 0$ galaxy stellar mass function and gas fractions of low-redshift groups and clusters (Kugel et al., 2023). The calibration forces getting the chosen calibration statistics right while allowing the inclusion of observational uncertainties, systematics, and biases.

A major benefit of the FLAMINGO simulation suite is that it includes multiple variants of baryonic feedback within a cosmologically representative volume. Baryonic feedback is one of the largest known systematic uncertainties in WL inferences. Baryonic feedback, specifically the accretion of supermassive black holes and supernova explosions, changes the matter distribution on (mildly) nonlinear scales, and they therefore impact WL estimates on this same scale. As the exact impact of these processes remains uncertain, it is useful to compare models with varying strengths while keeping all other parts of the model fixed. Additionally, whereas previous simulations, e.g., BAHAMAS (McCarthy et al., 2017), varied baryonic feedback by adjusting a single, unobservable, and unphysical subgrid parameter, FLAMINGO instead varies several parameters to shift the observables to which it was calibrated by a set amount. In this way, variations are identified by their distinct predictions for observables, allowing for more

straightforward interpretation and comparison and even enabling the application of observation-based priors to simulation predictions.

1.8 Baryonic impact on weak lensing

As WL probes the amount of matter, as quantified by Ω_m, and its distribution, as typically quantified by σ_8 or S_8, it is sensitive to feedback processes that distribute matter on (mildly) non-linear scales. Early studies quantified the enhancement due to the cooling of baryons using toy models (White, 2004), or by comparing the ratio of power spectra from DMO simulations to those that include baryonic physics (e.g., Jing et al., 2006; van Daalen et al., 2011; Chisari et al., 2018). As the WL 2-point statistics are modelled as a projection of the matter power spectrum (Eqs. 1.20 & 1.23), the impact of baryons on WL two-point statistics can be studied by comparing the signals using power spectra from DMO simulations to those generated from hydrodynamical simulations (Semboloni et al., 2011).

The impact of baryonic physics on HOWLS is typically much less understood. With the upcoming precise measurements from Stage IV WL surveys, studying these effects is both timely and necessary to exploit the data from these billion-dollar instruments to the fullest extent. As most of HOWLS lack a solid theoretical framework for modelling, other empirical methods must be used to quantify the impact of baryonic physics on HOWLS. The first, and most self-consistent, method is the direct generation of mock signals in simulations. By assuming certain survey properties (e.g., source redshift distribution, smoothing, PSF properties, and noise) and comparing signals measured from maps generated with and without baryons, one can estimate their impact. Although this method is the most self-consistent way to predict the impact of baryonic physics, it is also the most expensive. It has, therefore, so far not been possible to create enough simulations to directly marginalise over the impact of baryonic physics. A second commonly used method for resolving this issue is a halo model (Seljak, 2000; Smith et al., 2003; Asgari et al., 2023). A halo model is an analytical prescription that assumes all matter can be approximated as residing in haloes. In these models, the impact of baryons can also be approximated by free parameters (Mead et al., 2015, 2021), that can be marginalised over in a cosmological inference. However, these models, in particular their baryonic feedback, still need to be calibrated and validated using cosmological hydrodynamical simulations. The halo model is particularly useful for inferences of WL two-point statistics, as the halo distribution can be related to the power spectrum. However, for HOWLS, which typically require WL measurements at the map level, the approach does not work. A third method that can be used in this case is the baryonification method (Schneider & Teyssier, 2015; Aricò et al., 2021). These methods change the distribution of particles from much cheaper N-body simulations to mimic

1

the impact of baryons by moving individual particles according to a physical prescription. The outputs can then still be compared with those of the N-body simulation to estimate the impact of baryonic feedback (see, e.g., Weiss et al., 2019; Lu & Haiman, 2021). As for the halo model, these methods still need to be calibrated and validated using hydrodynamical simulations. These methods are typically highly flexible, with up to $\mathcal{O}(10)$ free parameters, and can therefore reproduce the outcomes of cosmological simulations (e.g., Pranjal R. et al., 2025). However, the models are typically relatively simple in terms of physical modelling and contain parameters for which it may be difficult to specify an observational prior. All these models so far also assume that the effects of cosmology and baryonic feedback are separable. As Stage IV surveys aim to achieve per cent-level precision on cosmological parameters, this coupling can likely not be ignored for at least their final data analyses (Elbers et al., 2025), which thus requires a more complex model that includes this coupling or the use of a suite of simulations that jointly vary cosmology and baryonic physics.

1.9 Matter power spectrum reconstruction

As demonstrated in Eqs. 1.20 & 1.23, the cosmic shear two-point statistics are modelled as projections of the matter power spectrum. Generally, cosmological inferences assume a cosmological model (typically ΛCDM) and constrain the best-fitting cosmological parameters. In the case of cosmic shear, these are primarily Ω_m and σ_8 or a combination of these. The resulting cosmology can be compared with that from surveys that exploit different physical mechanisms to construct a consistent picture of the Universe. When surveys do not yield consistent results, this may be due to systematic errors in either or both measurements, or to differences in the physical laws governing the scales and times they probe. Discrepancies may, for instance, be caused by the misestimation of known physical processes that impact a measurement (e.g., baryonic feedback), or by more complex physical scenarios (e.g., decaying or self-interacting DM), which is referred to as 'new physics'.

When tensions are observed, they may be a result of differences in cosmic time or scale, as shown in Figure 1.1. Therefore, instead of condensing the results into best-fit cosmological parameters, analyses may infer the shape of $P_\mathrm{m}(k,z)$ to gain insight into the causes of the observed discrepancies. For example, some processes, such as baryonic feedback and new physics components, leave known suppression imprints on the matter power spectrum (e.g., Preston et al., 2025; Schaller et al., 2025). Additionally, the method allows testing the internal consistency between the inferred cosmology and the best-fitting power spectrum.

This method has been used for more than two decades already (for early works, see e.g., Seljak, 1998; Pen et al., 2003), but has been popularised in the last few years in the WL community as Amon & Efstathiou (2022) suggested that

the origin of the S_8 tension between WL and *Planck* could be explained by strong suppression on (mildly) non-linear scales.

1.10 This thesis

This thesis studies the impact of baryonic feedback on WL observables using the FLAMINGO suite of cosmological hydrodynamical simulations and matter power spectrum reconstruction from Stage III cosmic shear data. The content can be split into two different parts. In Chapters 2 – 4, the cosmological potential and impact of baryonic feedback on different properties of WL peaks are studied using the FLAMINGO simulations. Chapters 5 & 6 carry out a matter power spectrum reconstruction using current state-of-the-art WL shear two-point correlation function data.

In Chapter 2, high-resolution ray-traced WL convergence maps are generated using the light-cone data of the FLAMINGO simulation suite. The maps are generated mimicking a *Euclid*-like survey. WL peak distributions are measured from these maps. We find that the peak-height distribution is more sensitive to realistic changes in cosmology than to changes in the baryonic feedback strength or mode of baryonic feedback. Nevertheless, the impact of baryonic feedback on WL peaks, as measured by Stage IV WL surveys, is non-negligible and should be carefully considered in their analyses.

In Chapter 3, we carry out a similar analysis as in Chapter 2, but we now focus on the redshift distribution of the same set of WL peaks. Using the FLAMINGO halo light-cone, we match high-valued WL peaks to individual haloes each. We illustrate that these high-valued peaks probe massive haloes with a completeness and purity competitive with current X-ray and SZ cluster abundance inferences. Additionally, we assign a redshift to each peak and take into account the expected photometric redshift uncertainty of *Euclid*. We find that the peak-redshift distribution not only contains valuable cosmological information but is also insensitive to the baryonic feedback strength, allowing us not only to achieve tighter and more robust cosmological constraints but also to calibrate baryonic feedback simultaneously while constraining cosmology.

In Chapter 4, we study the cosmological constraining power of the height, redshift, and angular clustering distributions of high-valued WL peaks using a new DMO hypercube, which varies the cosmology in a 10-dimensional space, including dark energy, massive neutrinos, and decaying dark matter. We find, in general, that of the considered WL peak properties, the redshift distribution provides the tightest constraints. Nevertheless, when all three statistics are combined, degeneracies are broken, yielding the tightest constraints.

In Chapter 5, we reconstruct the matter power spectrum using public data from Stage III WL surveys. For the matter power spectrum, we assume a double power-law in both the scale factor and the wavenumber. Using a Markov chain

1

Monte Carlo analysis, we find the best-fitting amplitude and evolution of the matter power spectrum, as preferred by the data. In general, we find that Stage III WL surveys are unable to provide tight constraints on the shape of the matter power spectrum, and most of their constraining power can be summarised by a single pivot point. Nonetheless, we find that KiDS and HSC show no evidence of evolution over time, in contradiction to ΛCDM predictions.

In Chapter 6, we carry out the same analysis as in Chapter 5, but using the new KiDS-Legacy data. Additionally, we use a more sophisticated deprojected regularisation approach to model the matter power spectrum. We find that the KiDS-Legacy data show improved internal consistency. The data also show a preference for small-scale suppression, consistent with the expectation of strong baryonic feedback.

References

Aarseth S. J., 1963, MNRAS, 126, 223
Aarseth S. J., 1966, MNRAS, 132, 35
Abbott T. M. C., et al., 2020, Phys. Rev. D, 102, 023509
Ahmad A., Cohen L., 1973, Journal of Computational Physics, 12, 389
Albrecht A., et al., 2006, arXiv e-prints, pp astro–ph/0609591
Allys E., Marchand T., Cardoso J. F., Villaescusa-Navarro F., Ho S., Mallat S., 2020, Phys. Rev. D, 102, 103506
Amon A., Efstathiou G., 2022, MNRAS, 516, 5355
Aricò G., Angulo R. E., Contreras S., Ondaro-Mallea L., Pellejero-Ibañez M., Zennaro M., 2021, MNRAS, 506, 4070
Asgari M., Mead A. J., Heymans C., 2023, The Open Journal of Astrophysics, 6, 39
Bacon D. J., Refregier A. R., Ellis R. S., 2000, MNRAS, 318, 625
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Barthelemy A., Codis S., Bernardeau F., 2021, MNRAS, 503, 5204
Benisty D., 2021, Physics of the Dark Universe, 31, 100766
Betancort-Rijo J., Patiri S. G., Prada F., Romano A. E., 2009, MNRAS, 400, 1835
Blandford R., Narayan R., 1986, ApJ, 310, 568
Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517
Bocquet S., et al., 2019, ApJ, 878, 55
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Boruah S. S., Rozo E., Fiedorowicz P., 2022, MNRAS, 516, 4111
Boughn S., Crittenden R., 2004, Nature, 427, 45
Broxterman J. C., Kuijken K., 2024, A&A, 692, A201
Broxterman J. C., et al., 2024, MNRAS, 529, 2309
Broxterman J. C., Schaller M., Hoekstra H., Schaye J., McGibbon R. J., Forouhar Moreno V. J., Kugel R., Elbers W., 2025, MNRAS, 538, 755
Cheng S., Ting Y.-S., Ménard B., Bruna J., 2020, MNRAS, 499, 5902
Chevallier M., Polarski D., 2001, International Journal of Modern Physics D, 10, 213
Chisari N. E., et al., 2018, MNRAS, 480, 3962
Chua K. T. E., Vogelsberger M., Pillepich A., Hernquist L., 2022, MNRAS, 515, 2681
Chudaykin A., Ivanov M. M., 2023, Phys. Rev. D, 107, 043518
Crain R. A., van de Voort F., 2023, ARA&A, 61, 473
DESI Collaboration et al., 2025a, J. Cosmology Astropart. Phys., 2025, 021
DESI Collaboration et al., 2025b, J. Cosmology Astropart. Phys., 2025, 012
DESI Collaboration et al., 2025c, J. Cosmology Astropart. Phys., 2025, 028
Dark Energy Survey and Kilo-Degree Survey Collaboration et al., 2023, The Open Journal of Astrophysics, 6, 36
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Davies C. T., Cautun M., Giblin B., Li B., Harnois-Déraps J., Cai Y.-C., 2022, MNRAS, 513, 4729

de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, Experimental Astronomy, 35, 25
Dietrich J. P., Hartlap J., 2010, MNRAS, 402, 1049
Dolag K., et al., 2025, arXiv e-prints, p. arXiv:2504.01061
Dyson F. W., Eddington A. S., Davidson C., 1920, Philosophical Transactions of the Royal Society of London Series A, 220, 291
Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241
Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1P
Einstein A., 1911, Annalen der Physik, 340, 898
Einstein A., 1916, Annalen der Physik, 354, 769
Einstein A., 1917, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, pp 142–152
Eisenstein D. J., et al., 2005, ApJ, 633, 560
Elbers W., et al., 2025, MNRAS, 537, 2160
Euclid Collaboration: Ajani V., Baldi M., Barthelemy A., et al., 2023, A&A, 675, A120
Euclid Collaboration: Castander F., Fosalba P., Stadel J., et al., 2025, A&A, 697, A5
Euclid Collaboration: Mellier Y., Abdurro'uf Acevedo Barroso J., et al., 2025, A&A, 697, A1
Euclid Collaboration: Vinciguerra S., Bouchè F., Martinet N., et al., 2025, A&A, submitted, p. arXiv:2510.04953
Feldbrugge J., van Engelen M., van de Weygaert R., Pranav P., Vegter G., 2019, J. Cosmology Astropart. Phys., 2019, 052
Forouhar Moreno V. J., Helly J., McGibbon R., Schaye J., Schaller M., Han J., Kugel R., Bahé Y. M., 2025, MNRAS,
Friedmann A., 1922, Zeitschrift fur Physik, 10, 377
Friedmann A., 1924, Zeitschrift fur Physik, 21, 326
García-García C., Ruiz-Zapatero J., Alonso D., Bellini E., Ferreira P. G., Mueller E.-M., Nicola A., Ruiz-Lapuente P., 2021, J. Cosmology Astropart. Phys., 2021, 030
Gatti M., et al., 2024, MNRAS, 527, L115
Ghirardini V., et al., 2024, A&A, 689, A298
Gilbank D. G., Gladders M. D., Yee H. K. C., Hsieh B. C., 2011, AJ, 141, 94
Hamana T., Takada M., Yoshida N., 2004, MNRAS, 350, 893
Henden N. A., Puchwein E., Shen S., Sijacki D., 2018, MNRAS, 479, 5385
Hernández-Aguayo C., et al., 2023, MNRAS, 524, 2556
Heydenreich S., Brück B., Harnois-Déraps J., 2021, A&A, 648, A74
Heymans C., et al., 2006, MNRAS, 368, 1323
Heymans C., et al., 2012, MNRAS, 427, 146
Heymans C., et al., 2021, A&A, 646, A140
Hildebrandt H., et al., 2017, MNRAS, 465, 1454
Hirata C. M., Mandelbaum R., Ishak M., Seljak U., Nichol R., Pimbblet K. A., Ross N. P., Wake D., 2007, MNRAS, 381, 1197
Hoekstra H., Jain B., 2008, Annual Review of Nuclear and Particle Science, 58, 99
Ishiyama T., et al., 2021, MNRAS, 506, 4210
Jain B., Van Waerbeke L., 2000, ApJ, 530, L1
Jing Y. P., Zhang P., Lin W. P., Gao L., Springel V., 2006, ApJ, 640, L119
Kaiser N., Wilson G., Luppino G. A., 2000, arXiv e-prints, pp astro–ph/0003338
Kamionkowski M., Riess A. G., 2023, Annual Review of Nuclear and Particle Science, 73, 153
Katz N., 1992, ApJ, 391, 502
Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
Kay S. T., Thomas P. A., Theuns T., 2003, MNRAS, 343, 608
Kazantzidis L., Perivolaropoulos L., 2018, Phys. Rev. D, 97, 103503
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Kilbinger M., et al., 2017, MNRAS, 472, 2126
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., Huffenberger K., 2012, Phys. Rev. D, 85, 103513
Kugel R., et al., 2023, MNRAS, 526, 6103
Kugel R., Schaye J., Schaller M., Forouhar Moreno V. J., McGibbon R. J., 2025, MNRAS, 537, 2179
LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201
Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
Leauthaud A., et al., 2007, ApJS, 172, 219
Lee M. E., Haiman Z., Pandey S., Genel S., 2026, ApJ, 996, 36
Lemaître G., 1931, MNRAS, 91, 483
Lewis A., Challinor A., 2006, Phys. Rep., 429, 1
Li Z., Liu X., Fan Z., 2023, MNRAS, 520, 6382
Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
Limber D. N., 1953, ApJ, 117, 134
Linder E. V., 2003, Phys. Rev. Lett., 90, 091301
Liu J., Haiman Z., 2016, Phys. Rev. D, 94, 043533
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Lu T., Haiman Z., 2021, MNRAS, 506, 3406
Lucas Makinen T., Heavens A., Porqueres N., Charnock T., Lapel A., Wandelt B. D., 2025, J. Cosmology Astropart. Phys., 2025, 095
Maartens R., Abdalla F. B., Jarvis M., Santos M. G., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 16, doi:10.22323/1.215.0016
Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, MNRAS, 242, 43
Marian L., Smith R. E., Bernstein G. M., 2009, ApJ, 698, L33
Marian L., Smith R. E., Hilbert S., Schneider P., 2013, MNRAS, 432, 1338
Maturi M., Angrick C., Pace F., Bartelmann M., 2010, A&A, 519, A23
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, MNRAS, 454, 1958
Mead A. J., Brieden S., Tröster T., Heymans C., 2021, MNRAS, 502, 1401
Miyazaki S., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460Z, doi:10.1117/12.926844
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Palanque-Delabrouille N., et al., 2015, J. Cosmology Astropart. Phys., 2015, 011
Pantos I., Perivolaropoulos L., 2026, arXiv e-prints, p. arXiv:2602.12238
Peebles P. J. E., 1982, ApJ, 263, L1
Pen U.-L., Lu T., van Waerbeke L., Mellier Y., 2003, MNRAS, 346, 994
Perlmutter S., et al., 1999, ApJ, 517, 565
Philcox O. H. E., Ivanov M. M., 2022, Phys. Rev. D, 105, 043517
Pizzati E., et al., 2024, MNRAS, 534, 3155
Planck Collaboration et al., 2020a, A&A, 641, A6
Planck Collaboration et al., 2020b, A&A, 641, A8
Porqueres N., Heavens A., Mortlock D., Lavaux G., 2022, MNRAS, 509, 3194
Pranjal R. S., Pandey S., Anbajagane D., Krause E., Dolag K., 2025, arXiv e-prints, p. arXiv:2507.13317
Preston C., Rogers K. K., Amon A., Efstathiou G., 2025, MNRAS,
Reid B. A., et al., 2010, MNRAS, 404, 60
Riess A. G., et al., 1998, AJ, 116, 1009
Robertson H. P., 1935, ApJ, 82, 284
Robertson H. P., 1936, ApJ, 83, 257
Schaller M., Schaye J., Kugel R., Broxterman J. C., van Daalen M. P., 2025, MNRAS, 539, 1337
Schaye J., 2004, ApJ, 609, 667
Schaye J., et al., 2023, MNRAS, 526, 4978
Schneider P., 1985, A&A, 143, 413
Schneider A., Teyssier R., 2015, J. Cosmology Astropart. Phys., 2015, 049

Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002, A&A, 396, 1
Schneider P., Eifler T., Krause E., 2010, A&A, 520, A116
Secco L. F., et al., 2022, Phys. Rev. D, 105, 023515
Seljak U., 1998, ApJ, 506, 64
Seljak U., 2000, MNRAS, 318, 203
Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 2020
Silvestre-Rosello E., Porth L., Schneider P., Linke L., Krueger J., Grandis S., Oel J., 2025, arXiv e-prints, p. arXiv:2509.07973
Smith R. E., et al., 2003, MNRAS, 341, 1311
Spergel D. N., et al., 2003, ApJS, 148, 175
Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Springel V., et al., 2018, MNRAS, 475, 676
Stiskalek R., 2025, arXiv e-prints, p. arXiv:2509.20235
Stölzner B., et al., 2025, A&A, 702, A169
Takada M., Jain B., 2003, MNRAS, 344, 857
Taruya A., Takada M., Hamana T., Kayo I., Futamase T., 2002, ApJ, 571, 638
Tegmark M., et al., 2004, Phys. Rev. D, 69, 103501
The Dark Energy Survey Collaboration 2005, arXiv e-prints, pp astro–ph/0510346
Tröster T., et al., 2020, A&A, 633, L10
Van Waerbeke L., et al., 2000, A&A, 358, 30
Van Waerbeke L., et al., 2001, A&A, 374, 757
Van Waerbeke L., et al., 2013, MNRAS, 433, 3373
van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, MNRAS, 415, 3649
Velliscig M., et al., 2015, MNRAS, 453, 721
Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42
von Hoerner S., 1960, Z. Astrophys., 50, 184
von Soldner J. G., 1804, Berliner Astron. Jahrb., 29, 161
Walker A. G., 1937, Proceedings of the London Mathematical Society, 42, 90
Wang S., Haiman Z., May M., 2009, ApJ, 691, 547
Wang Q., et al., 2025, arXiv e-prints, p. arXiv:2510.27291
Weinberg S., 1989, Reviews of Modern Physics, 61, 1
Weiss A. J., Schneider A., Sgier R., Kacprzak T., Amara A., Refregier A., 2019, J. Cosmology Astropart. Phys., 2019, 011
White M., 2004, Astroparticle Physics, 22, 211
White S. D. M., Efstathiou G., Frenk C. S., 1993, MNRAS, 262, 1023
White M., et al., 2022, J. Cosmology Astropart. Phys., 2022, 007
Wiersma R. P. C., Schaye J., Smith B. D., 2009, MNRAS, 393, 99
Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, Nature, 405, 143
Wittman D. M., et al., 2002, in Tyson J. A., Wolff S., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4836, Survey and Other Telescope Technologies and Discoveries. pp 73–82 (arXiv:astro-ph/0210118), doi:10.1117/12.457348
Wolf C., Meisenheimer K., Rix H. W., Borch A., Dye S., Kleinheinrich M., 2003, A&A, 401, 73
Wright A. H., et al., 2025, A&A, 703, A158
Yang X., Kratochvil J. M., Wang S., Lim E. A., Haiman Z., May M., 2011, Phys. Rev. D, 84, 043529
Zwicky F., 1933, Helvetica Physica Acta, 6, 110
Zwicky F., 1937, ApJ, 86, 217

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

Authors: **Jeger C. Broxterman**, Matthieu Schaller, Joop Schaye, Henk Hoekstra, Konrad Kuijken, John C. Helly, Roi Kugel, Joey Braspenning, Willem Elbers, Carlos S. Frenk, Juliana Kwan, Ian G. McCarthy, Jaime Salcido, Marcel P. van Daalen, and Bert Vandenbroucke

Published in Monthly Notices of the Royal Astronomical Society, 529, 2309 (2024)

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 AGN feedback. We show that the differences induced by realistic baryonic feedback prescriptions (typically 5−30% for $\kappa = 0.1-0.4$) are smaller than those induced by reasonable variations in cosmological parameters ($20-60$% for $\kappa = 0.1-0.4$) but must be modeled 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.

2.1 Introduction

Over the last few decades, the spatially flat ΛCDM 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 et al., 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 ΛCDM 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); 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 et al., 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, 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. Asgari et al., 2021; Amon et al., 2022; Hamana et al., 2020). Astrophysical feedback processes (e.g. supernova explosions, and active galactic nuclei (AGN)) 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 $k \gtrapprox 0.1\ h\,\mathrm{Mpc}^{-1}$ and enhances the power on even smaller scales (e.g. van Daalen et al., 2011, 2020; Chisari et al., 2018; Schneider et al., 2020; Schaye et al., 2023; Salcido 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, 2013; Gouin et al., 2019; Weiss et al., 2019).

Two-point statistics encapsulate all cosmological information in the underly-

ing field if it can be described as a Gaussian random field (GRF). 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 et al., 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 et al., 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, for example by using a halo model (e.g. Sabyr et al., 2022; Asgari et al., 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 2-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 to 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 weak lensing statistics. The first approach, which we adopt, uses full-hydrodynamical simulations employing sub-grid 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 SZ. The second is to use N-body simu-

lations to model the evolution of dark matter, 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 to 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 (Schaye et al., 2023; Kugel 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. 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 to 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 2.2 we introduce the relevant WL and ray-tracing theory. Section 2.3 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 2.4 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 stellar mass function. In this way, we are able to understand the relation between stronger feedback, leading to lower gas fractions in clusters or lower galaxy stellar mass functions, and the observed WL peak distribution. We then compare the different AGN subgrid models and the different cosmology variations in Section 2.5, where we also look at the separability of the baryonic and cosmological effects. We conclude by comparing the differences induced by baryonic feedback variations to those due to changes in cosmology. Our main results are summarized in Section 2.6.

2.2 Weak Lensing 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 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, $\boldsymbol{\alpha}$, which at each angular position $\boldsymbol{\theta}$ 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

$$\mathrm{d}\boldsymbol{\alpha} = -\frac{2}{c^2}\nabla_\perp \Phi \,\mathrm{d}\chi, \tag{2.1}$$

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,

$$\boldsymbol{\beta}(\boldsymbol{\theta}, \chi) = \boldsymbol{\beta}(\boldsymbol{\theta}, 0) - \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{\chi - \chi'}{\chi\chi'} \nabla_\perp \Phi(\boldsymbol{\beta}(\boldsymbol{\theta}, \chi'), \chi'), \tag{2.2}$$

where $\boldsymbol{\beta}(\boldsymbol{\theta}, 0) = \boldsymbol{\theta}$ 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

$$A_{ij}(\boldsymbol{\theta}, \chi) \equiv \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \delta_{ij} - \frac{2}{c^2} \int_0^{\chi} \mathrm{d}\chi' \, \frac{\chi - \chi'}{\chi \chi'} \, \partial_i \partial_j \Phi(\boldsymbol{\beta}(\boldsymbol{\theta}, \chi'), \chi'), \tag{2.3}$$

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

$$\begin{aligned} A &\equiv \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} \\ &\approx \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2+\omega \\ -\gamma_2-\omega & 1-\kappa+\gamma_1 \end{pmatrix}, \end{aligned} \tag{2.4}$$

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 et al. (2000). κ is the lensing convergence, which we will use to quantify the WL strength and $\gamma = \gamma_1 + \mathrm{i}\gamma_2$ 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. $\boldsymbol{\beta}(\boldsymbol{\theta}, \chi) \cong \boldsymbol{\beta}(\boldsymbol{\theta})$). 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 galaxygalaxy lensing shear profiles (Simon & Hilbert, 2018), the CMB lensing bispectrum (Pratten & Lewis, 2016), and higher order moments of convergence profiles, where Petri et al. (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 analyze the maps, the theoretical equations 2.2 and 2.3 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 (eqs. 2.2 and 2.3) can be expressed as a sum over a discrete number of lensing planes such that at the n-th plane they are given by

$$\boldsymbol{\beta}^n = \boldsymbol{\beta}^0 - \sum_{m=0}^{n-1} \frac{\chi^n - \chi^m}{\chi^n} \boldsymbol{\alpha}^m, \tag{2.5}$$

and

$$A_{ij}^n = \delta_{ij} - \sum_{m=0}^{n-1} \frac{\chi^n - \chi^m}{\chi^n} U_{ik}^m A_{kj}^m, \tag{2.6}$$

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 $\boldsymbol{\theta}$. Here, we defined the shear matrix U such that the deflection field ($\boldsymbol{\alpha}^m$) and shear matrix equal the first- ($\boldsymbol{\alpha}^n(\boldsymbol{\theta}) = \nabla_\perp \psi^n(\boldsymbol{\theta})$) 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

$$\psi^n \equiv \frac{2}{\chi^n c^2} \int_{\chi_{\min}}^{\chi_{\max}} \mathrm{d}\chi \, \Phi, \tag{2.7}$$

where $\chi_{\min}$ and $\chi_{\max}$ are the comoving distances to the beginning and end of the shell, respectively. In principle, using Equations 2.5 and 2.6, 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 Roberston-Walker metric and is given by (Schneider, 2016):

$$\begin{aligned} A_{ij}^{n+1} = &\left(1 - \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}}\right) A_{ij}^{n-1} \\ &+ \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}} A_{ij}^n - \frac{\chi^{n+1} - \chi^n}{\chi^{n+1}} U_{ik}^n A_{kj}^n, \end{aligned} \tag{2.8}$$

with the initialisation 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 ($\boldsymbol{\beta}^{n+1}$) can be evaluated using the position at the current plane ($\boldsymbol{\beta}^n$) and the previous plane ($\boldsymbol{\beta}^{n-1}$),

$$\begin{aligned} \boldsymbol{\beta}^{n+1} = &\left(1 - \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}}\right) \boldsymbol{\beta}^{n-1} \\ &+ \frac{\chi^n}{\chi^{n+1}} \frac{\chi^{n+1} - \chi^{n-1}}{\chi^n - \chi^{n-1}} \boldsymbol{\beta}^n - \frac{\chi^{n+1} - \chi^n}{\chi^{n+1}} \boldsymbol{\alpha}^n, \end{aligned} \tag{2.9}$$

where $\beta^{-1} = \beta^0 = \theta$ are the initial angular positions of the rays. In Section 2.3.4 we will explain how the FLAMINGO lightcones can be related to the deflection field and shear matrix in Equations 2.8 & 2.9. The equations can then be combined with Equation 2.4 to estimate the WL convergence, κ.

2

2.3 Methods

2.3.1 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 large-scale structure 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., 2024) with the SPHENIX smoothed particle hydrodynamics (SPH) implementation (Borrow et al., 2022). Neutrinos are modeled 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 (Ploeckinger & Schaye, 2020), star formation (Schaye & Dalla Vecchia, 2008), and time-dependent stellar mass loss as described by Wiersma et al. (2009). Supernova (SN) and stellar feedback are implemented kinetically (Dalla Vecchia & Schaye, 2008) by kicking neighboring 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 onto supermassive black holes 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 stellar mass function (SMF) and gas fractions in low$-z$ clusters. Comparing sets of simulations calibrated to different observables allows for more instructive comparisons than comparing simulations that differ in specific subgrid parameters. Additionally, expected observational biases were included

Table 2.1: Characteristics of the 4 simulation variations that differ only in numerical resolution and box size. The columns indicate the simulation identifier; the box size length in comoving Gpc, 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 $\mathrm{M}_\odot$	m_CDM $\mathrm{M}_\odot$
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	1.07×10^9	5.65×10^9

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 supernova and AGN feedback, namely the fraction of stellar energy feedback that couples to the ISM (f_SN), the target wind velocity for supernova feedback (Δv_SN), the black hole (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 4 different runs varying the mass resolution and box size. The details of these runs are listed in Table 2.1. 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 (5040^3 for dark matter (DM) and baryons and 2800^3 for massive neutrinos) particles with a baryonic particle mass of $1.07 \times 10^9\,\mathrm{M}_\odot$, 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 postfix '_DMO'. The cosmology of these runs was taken from the Dark Energy Survey year three '3×2pt+ All Ext.' ΛCDM cosmology (Abbott et al., 2022), indicated as 'D3A' in Table 2.3. The initial conditions were generated using MonofonIC (Hahn et al., 2020; Michaux et al., 2021) using 3-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 $z = 31$.

2.3.2 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 2.2. At fixed cosmology, 8 runs were calibrated to different galaxy stellar mass functions (M*) and/or gas fractions in clusters ($f_{\rm gas}$) and/or differ in overall AGN subgrid feedback prescription. The subgrid parameter values have been chosen such that the stellar mass function and/or gas fraction in clusters within the simulation are a set number of standard deviations from the fiducial model (ΔM* and $\Delta f_{\rm gas}$, respectively), as indicated in the second and third columns of Table 2.2. 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 weak lensing 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). 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 4 runs varying only the gas fractions in clusters, whose identifiers are given by 'fgas±nσ', where n is the number of standard deviations by which the gas fraction is varied. One run was calibrated to match the galaxy stellar mass function shifted to a lower mass by 1σ (M*−σ). The M*−σ_fgas−4σ 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_fgas−4σ', where for the latter the target gas fraction in clusters is reduced. The Jet models can be used to assess the sensitivity of observables to variations in subgrid models calibrated to the same observables.

At fixed baryonic calibration, there are 4 cosmology variations. The different cosmologies and their cosmological parameters are listed in Table 2.3. In addition to the fiducial D3A cosmology that was used for all the different baryonic physics runs, there are 3 variations based on cosmic microwave background (CMB) measurement from Planck. The first CMB cosmology is the best-fit Planck Collaboration et al. (2020) ΛCDM cosmology with $\sum m_\nu = 0.06$ eV (Planck). The other two Planck cosmology variations include heavier neutrinos (3 species each with $m_\nu c^2 = 0.08$ eV), one in which the other parameters are changed according to their best-fit values within the Planck MCMC chains (PlanckNu0p24Var), and one in which the other parameter values are fixed and only $\Omega_{\rm CDM}$ is adjusted to keep $\Omega_{\rm m}$ 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.

Table 2.2: 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 stellar mass function (M*) and gas fractions in clusters ($f_{\rm gas}$) in the simulation are shifted compared to 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 2.3) and with the same initial conditions.

Identifier	ΔM* σ	$\Delta f_{\rm gas}$ σ	AGN mode
L1_m9	0	0	thermal
fgas+2σ	0	+2	thermal
fgas−2σ	0	-2	thermal
fgas−4σ	0	-4	thermal
fgas−8σ	0	-8	thermal
M*−σ	-1	0	thermal
M*−σ_fgas−4σ	-1	-4	thermal
Jet	0	0	jet
Jet_fgas−4σ	0	-4	jet

Table 2.3: The cosmological parameter values used in the different FLAMINGO simulations. The columns indicate the cosmology identifier; the dimensionless Hubble constant, $h = H_0/(100$ km/s/Mpc); the total matter density parameter, Ω_m; the baryonic matter density parameter, Ω_b; the cosmological constant density parameter, Ω_Λ; the neutrino mater 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 parameterized as the r.m.s. mass density fluctuation in spheres of radius 8 h^{-1} Mpc at $z = 0$, σ_8, and the amplitude of the initial power spectrum parametrized as $S_8 \equiv \sigma_8 \sqrt{\Omega_\mathrm{m}/0.3}$.

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

2.3.3 Lightcones

To construct the WL convergence maps, we use the mass maps of the FLAMINGO lightcones. The lightcones correspond to virtual observers within the FLAMINGO simulation suite. They have been constructed by recording particles crossing the past lightcone of a virtual observer. The lightcones 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 $N_{\rm side} = 16384$ (with a number of pixels of $N_{\rm pix} = 12N_{\rm side}^2$), corresponding to an angular resolution of 0.21 arcmin. For the details of the lightcone 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 2.3.4), the box is replicated to reach this distance. For the L2p8 and L1 boxes, this requires 5 and 12 additional replications, respectively. In Appendix 2.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 lightcone 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 lightcone 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, 8 observers have been placed at the coordinates $(\pm L/4, \pm L/4, \pm L/4)$, where L is the simulation box size. These lightcones have 68 shells up to $z = 5$ but we only use the first 60 lightcones to facilitate a direct comparison with the L1 lightcones, as the first 60 shells have the same redshift spacing as the lightcones of the L1 runs. The mean shell thickness is ≈ 110 Mpc. Zorrilla Matilla et al. (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 8 different lightcones to quantify the impact of cosmic variance on our analysis. Additionally, we carry out numerical convergence tests for simulation box size and resolution.

2.3.4 Full-sky WL convergence maps

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 $N_{\rm side} = 8192$. 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 lightcone maps. We therefore limit our analysis to downsampled maps with $N_{\rm side} = 8192$, 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 2.3.5. At each of the lightcone shells, we determine the lensing potential (given by Equation 2.7) using the solution of the two-dimensional Poisson equation which we define as the convergence at plane n (K^n) as

$$K^n(\boldsymbol{\theta}) \equiv \frac{1}{2}\nabla^2\psi^n = \frac{3H_0^2\Omega_{\rm m}}{2c^2}\,\chi^n\,(1+z^n)\,\Delta\chi^n\,\delta^n(\boldsymbol{\theta}), \tag{2.10}$$

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

$$\delta^n(\boldsymbol{\theta}) = \frac{\Sigma^n(\boldsymbol{\theta}) - \overline{\Sigma^n}}{\overline{\Sigma^n}}, \tag{2.11}$$

where $\Sigma^n(\boldsymbol{\theta})$ is the surface density at position $\boldsymbol{\theta}$ and $\overline{\Sigma^n}$ 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 lightcones, 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 ($K^n_{\ell m}$) and lensing potential ($\psi^n_{\ell m}$) as

$$\psi^n_{\ell m} = \frac{-2K^n_{\ell m}}{\ell(\ell+1)}, \tag{2.12}$$

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 centers. Since the photons are deflected here, their paths will, in general, not pass through the center 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 et al. (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 2.8 & 2.9), we determine the lensing potential using Equation 2.12. 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).

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 2.4. To mimic a signal corresponding to a Stage IV survey, we incorporate a source redshift distribution ($n(z)$) as

$$\kappa(\theta) = \int_0^{\chi_{\rm hor}} \mathrm{d}\chi\, n(z(\chi))\, \kappa(\theta, \chi), \tag{2.13}$$

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 $\chi_{\rm hor} = \chi(z = 3)$. Here, we use a simple *Euclid* mock forecast given by (Euclid Collaboration et al., 2020):

$$n(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right], \tag{2.14}$$

with $z_0 = 0.9/\sqrt{2}$. The function is normalized such that $\int n(z)\,\mathrm{d}z = 1$, and relates to the comoving distance as $n(z)\,\mathrm{d}z = n(\chi)\,\mathrm{d}\chi$. 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. 2.1. 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.

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

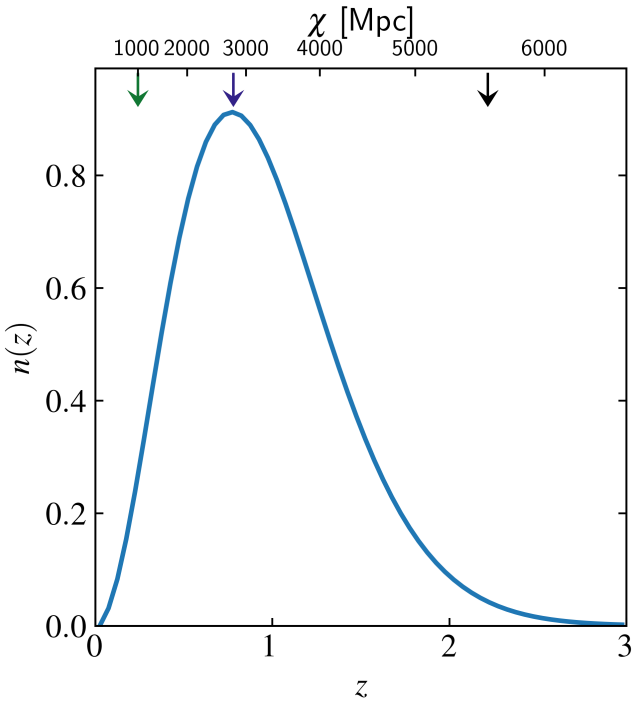


Figure 2.1: *Euclid*-like source redshift distribution used in the construction of the WL convergence maps. The top axis shows the comoving distance for the fiducial D3A cosmology. The green, blue, and black arrows indicate the box size of the L1 (1 cGpc), L2p8 (2.8 cGpc), and L5p6 (5.6 cGpc) variations, respectively.

σ (Kaiser & Squires, 1993; Lu & Haiman, 2021):

$$\mathcal{N}\left\{\mu = 0, \sigma = \frac{\sigma_\epsilon}{\sqrt{2 n_{\rm gal} A_{\rm pix}}}\right\}, \tag{2.15}$$

where σ_ϵ is the rms intrinsic ellipticity of source galaxies, $n_{\rm gal}$ is the source number density on the sky and $A_{\rm pix}$ is the pixel area. In our case, we choose $\sigma_\epsilon = 0.26$ and $n_{\rm gal} = 30$ 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* (*HST*) images with similar photometric properties as the expected *Euclid* images and is therefore commonly used to model the observed galaxy shape noise (Schrabback et al., 2018; Euclid Collaboration et al., 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.

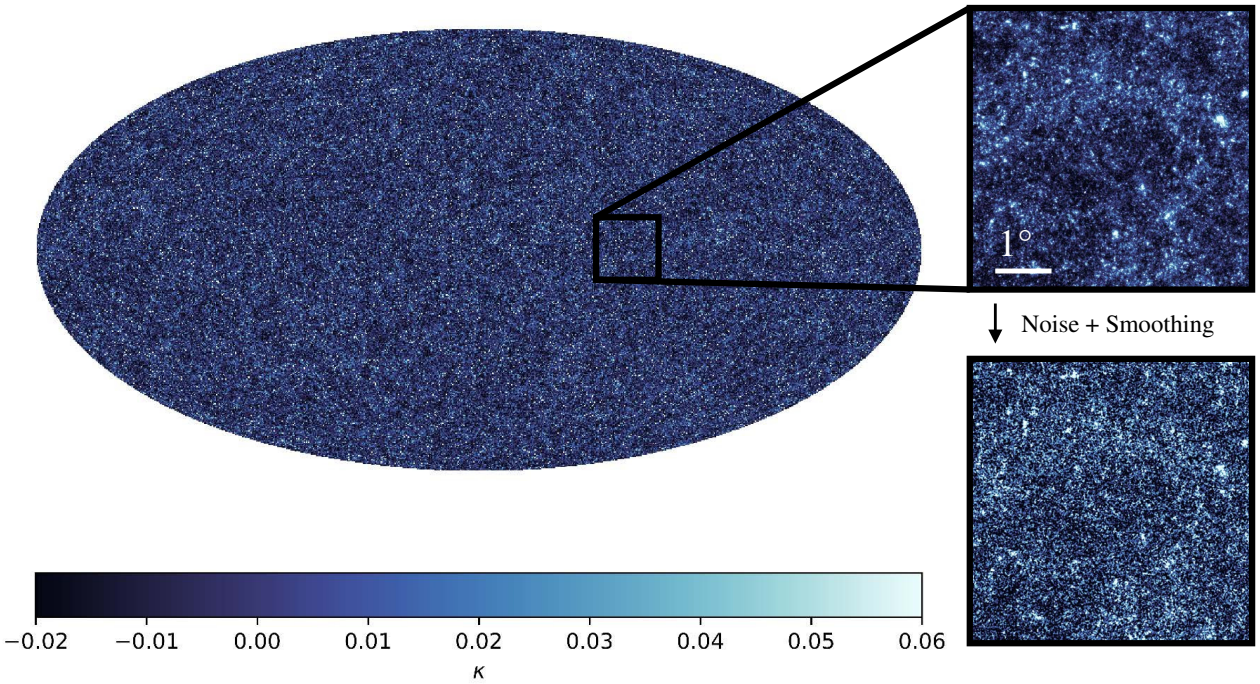


Figure 2.2: Ray-traced full-sky weak lensing convergence (κ) map (left) and zoom-ins of a 5×5 deg^2 region for a *Euclid*-like source redshift distribution as measured by a virtual observer in the hydrodynamical L2p8_m9 run. The full-sky map and top zoom-in correspond to a WL convergence map without 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.

Fig. 2.2 shows a full-sky WL convergence map corresponding to an observer in the L2p8_m9 simulation run. Here, the full-sky map and top 5×5 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.

2

Second-order effects

Our analysis is somewhat idealized as we ignore several second-order effects such as lens-lens coupling (Bernardeau et al., 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% of the signal on arc minute scales. However, in our analysis, we assume that the reconstruction of the signal takes into account the reduced shear correction, which can, for example, be done using a quickly converging iteration (Bradač et al., 2005).

2.3.5 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. 2.3. We compare to the prediction from HALOFIT (Takahashi et al., 2020) using the CLASS code (Blas et al., 2011). HALOFIT is a commonly-used fitting formula to the (non-linear) three-dimensional 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):

$$\mathcal{C}(\ell) = \int_0^{\chi_\mathrm{hor}} \mathrm{d}\chi \, \frac{W^2(\chi)}{\chi^2} \, P_\mathrm{m}\left(\frac{\ell+1/2}{\chi}, z(\chi)\right), \tag{2.16}$$

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

$$W(\chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2} \, \chi \, (1+z(\chi)) \int_\chi^{\chi_\mathrm{hor}} \mathrm{d}\chi' \, n(\chi') \, \frac{\chi'-\chi}{\chi'}. \tag{2.17}$$

The HALOFIT prediction is given by the dash-dotted green curve in Fig. 2.3. 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 gridpoint (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. (2024), 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 $N_{\rm side} = 4096$, 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 4 nearest neighbors on the two-dimensional 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 $N_{\rm side} = 4096$. 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 to NGP interpolation. At $\ell = 2000$ ($\theta \approx 5$ arcmin), the difference between the interpolation strategies is 1% 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 SNR < 6 that agrees to within 5% with codes that either use higher resolution equidistant cylindrical projection pixelization or determine the convergence signal on the fly (Muciaccia et al., 1997; Barreira et al., 2017; Fabbian et al., 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 to 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, 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

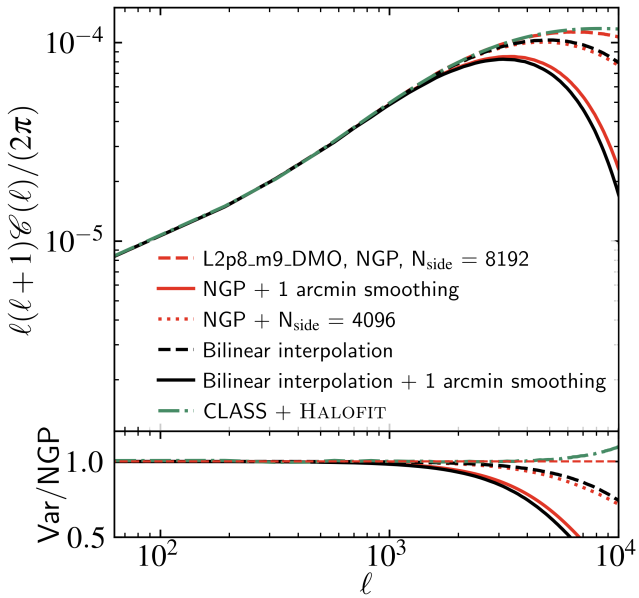


Figure 2.3: Top: Weak lensing convergence angular power spectrum for variations of smoothing, angular resolution, and interpolation strategy for a single observer in the L2p8_m9_DMO run and the theoretical non-linear prediction from CLASS + HALOFIT. Bottom: ratios of the different curves in the top panel and the one for nearest gridpoint (NGP) interpolation at $N_{\rm side} = 8192$ (dashed red curve). The 1 arcmin smoothing dominates over any suppression induced by pixelation or ray tracing but differences of a few per cent are still present between different procedures. We use bilinear interpolation at $N_{\rm side} = 8192$ with 1 arcmin smoothing (solid black curve) as our fiducial map construction method.

discretization of the HEALPix grid.

2.4 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 8 closest pixels on the HEALPix grid. We use the number density of peaks to study the peak distribu-

tions. 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].

2.4.1 Numerical convergence

As discussed in Section 2.3.2, 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. 2.4 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 8 observers in L2p8_m9_DMO (blue). The high-resolution run has a 64× higher mass resolution than the low-resolution run. The top axis of the figure shows the SNR = κ/σ, where σ is the standard deviation of the smoothed galaxy shape noise map (Equation 2.15). 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) % 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 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%. However, the number of peaks in this regime also drops and the distribution thus becomes dominated by Poisson noise, as shown by the gray-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

[1] Although 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]

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.

2.4.2 Cosmic variance

Using the 8 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. 2.5 shows the number density of peaks for each of the 8 lightcones. The bottom panel shows their ratios with their mean. The standard deviation of the ratio is indicated by the gray shaded region. For positive WL convergence values up to $\kappa = 0.1(0.2)$, the distributions agree to within 0.2 (1)% precision. Up to $\kappa \approx 0.4$, the difference is not greater than 10%, 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 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 to the Poisson uncertainty in Fig. 2.4, we see that over the entire κ

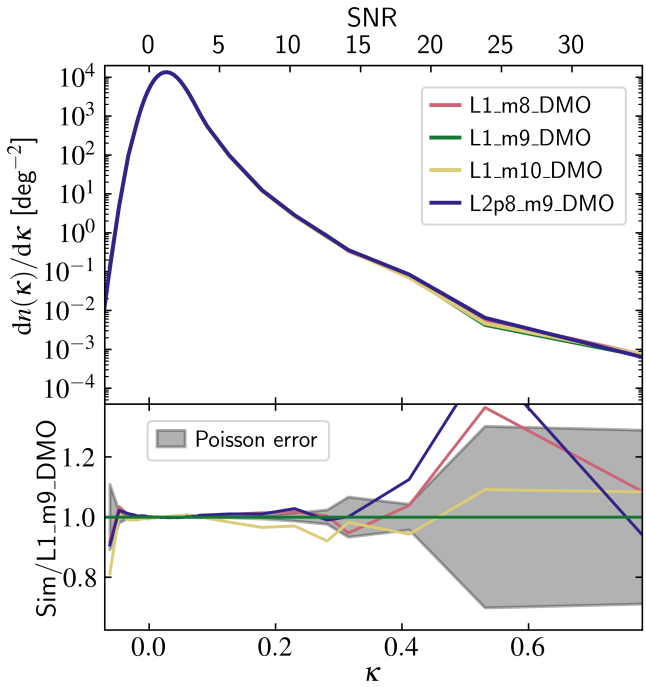


Figure 2.4: Top: number density of WL peaks for the observers in the L1_m8, L1_m9, L1_m10, and L2p8_m9 DMO runs, which differ in terms of resolution and/or box size. The L2p8_m9_DMO signal is the mean of the 8 observers in the box. The SNR = κ/σ (top x-axis) is computed from the standard deviation of a smoothed noise realization. Bottom: 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.

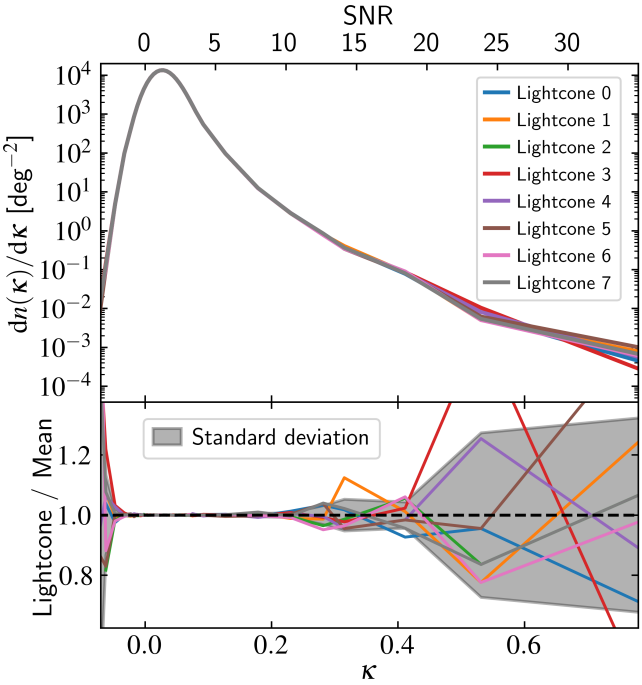


Figure 2.5: Top: number density of WL peaks for 8 different virtual observers in the L2p8_m9_DMO simulation. The observers were placed at the coordinates $(\pm L/4, \pm L/4, \pm L/4)$, where $L = 2.8$ cGpc is the simulation box size. Bottom: Ratio with the mean of the 8 observers. The gray shading shows the standard deviation of the 8 observers. Over the entire range of κ, the cosmic variance is comparable to 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.

range, the two are of similar magnitude.

2.4.3 Comparison of fiducial dark matter only 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 8 virtual observers in the L2p8_m9 and L2p8_m9_DMO runs. Their mean number density of peak distributions are shown in the top panel of Fig. 2.6 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 to 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 (≈ 15000 deg^2; Laureijs et al., 2011) and the KiDS footprint (≈ 1550 deg^2; 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 2.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 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 2.4.1 & 2.4.2. The gray-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 to the DMO run (Stanek et al., 2009; Cui et al., 2012). Debackere et al. (2022) compared the halo masses of matched haloes in BAHAMAS and its DMO counterpart and showed that for haloes of mass $M_{200\mathrm{m}} \approx [10^{13} - 10^{14}]\mathrm{M}_\odot$, the halo mass in the hydrodynamical simulation is $\sim 10\%$ lower than in the DMO run. As the WL convergence signal is directly proportional to the overdensity (Equation 2.10) 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 ($M_{200\mathrm{m}} \gtrsim 10^{14}$ 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 to the DMO counterpart, the HMF is suppressed in the intermediate

halo mass regime ($M_{200\mathrm{m}} \approx [10^{13} - 10^{14}]\,\mathrm{M}_\odot$) by 10-20%, with larger differences for less massive haloes, but for the most massive haloes ($M_{200\mathrm{m}} \approx 10^{15}\,\mathrm{M}_\odot$) 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 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 et al. (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 to the ones extracted by Ferlito et al. (2023) from the Millenium-TNG 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%, and that the suppression decreases

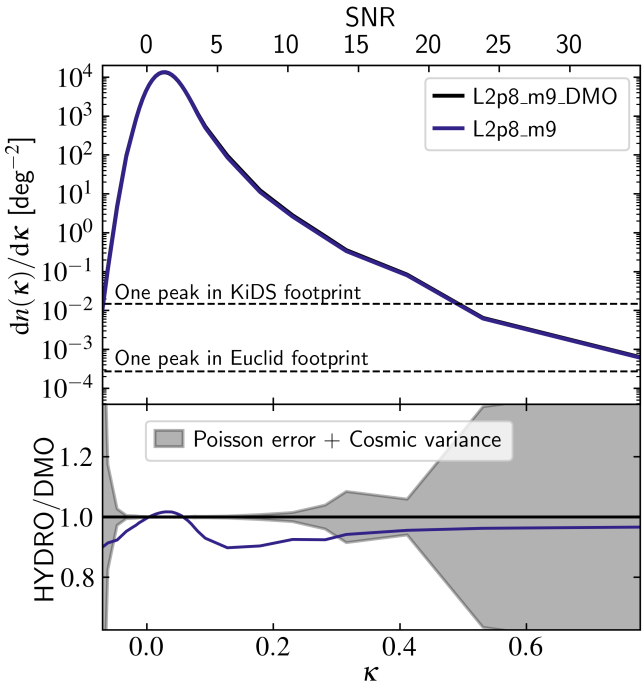


Figure 2.6: Top: mean number density of WL peaks measured by the 8 virtual observers in the DMO and cosmological hydrodynamical L2p8_m9 runs with the same cosmology and initial phases. The observers are located at the same positions in the simulations and we apply the same random rotations to their light-cone 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: ratio of the mean number density of peaks for the observers in the hydrodynamical run to the mean in the DMO run. The gray 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% suppression of the number density of WL peaks. The suppression due to baryons decreases almost monotonically for larger κ.

for larger κ values. 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.

2.4.4 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

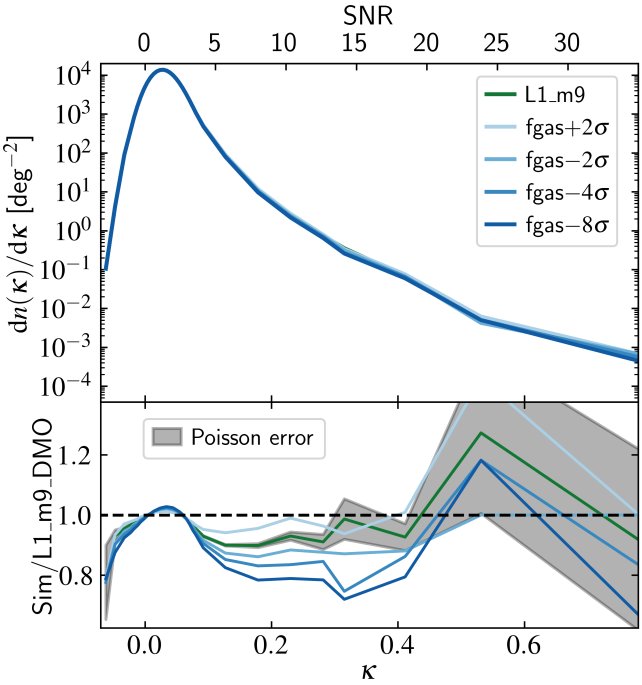


Figure 2.7: As Fig. 2.6 but comparing the runs calibrated to different gas fractions in cluster, as indicated by fgas−nσ where n is the change in the number of standard deviations compared to the fiducial L1_m9 (green curve) model, to L1_m9_DMO. Models with stronger feedback, indicated by an increasingly darker blue color, have progressively stronger suppressed number densities of WL peaks.

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 4 subgrid parameters, as described in Section 2.3.2, which were chosen to change the resulting gas fraction ($f_{\rm gas}$) in clusters by a set amount. The number density of peaks for the runs with different gas fractions are shown in Fig. 2.7. Again, at $\kappa < 0.1$, the differences between the distributions are washed out due to the applied smoothing and noise.

In general, compared to 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. 2.7 shows that the models with lower gas fractions, which are indicated by an increasingly darker blue color, 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 centers 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 $10^{13}\,\mathrm{M}_\odot \lesssim M_{200\mathrm{m}} \lesssim 5\times10^{14}\,\mathrm{M}_\odot$ 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. 2.8. 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%) at $\kappa = 0.1$, for model Jet it increases to $\approx 20\%$ at $\kappa = 0.4$. In contrast, the fgas−4σ (blue) and Jet_fgas−4σ (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 $\approx 10\%$ 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 $10^{13}\,\mathrm{M}_\odot \lesssim M_{200\mathrm{m}} \lesssim 10^{14}\,\mathrm{M}_\odot$ the thermal and jet models calibrated to the fiducial gas fraction show 5-20% differences, with L1_m9 having a lower HMF, whereas the fgas−4σ models do not differ by more than 10% from each other. At larger halo masses, up to $M_{200\mathrm{m}} \approx 5\times10^{14}\,\mathrm{M}_\odot$, the difference between L1_m9 and Jet vanishes whereas the difference between the fgas−4σ and Jet_fgas−4σ models remains of similar magnitude. Possibly, the peaks we measure primarily originate from haloes with $M_{200\mathrm{m}} > 10^{14}\,\mathrm{M}_\odot$, as the difference between the fgas−4σ models is larger in that regime. Liu &

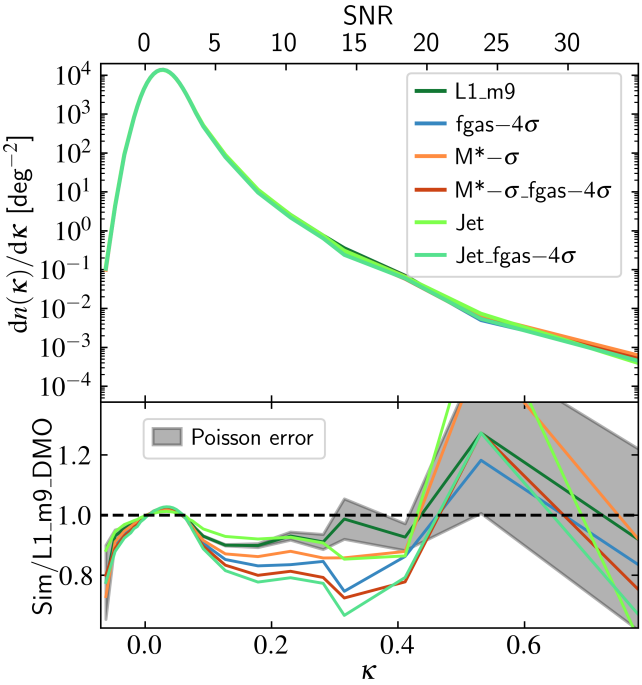


Figure 2.8: As Fig. 2.6 but comparing the baryonic feedback models varying the galaxy stellar mass function (M*), AGN feedback prescription (Jet), gas fraction in clusters (fgas), or a combination of these, to L1_m9_DMO. Calibration to lower galaxy stellar mass functions leads to suppressed number densities of WL peaks. Variations in AGN subgrid prescriptions can lead to differences in WL peak counts not captured by the gas fraction in clusters.

Haiman (2016) have used a halo model to study the origin of WL convergence peaks in the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) and found that the highest valued peaks are caused by a single massive halo of mass $M_{\rm vir} \approx 10^{15}\,{\rm M}_{\odot}$. 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 stellar mass functions. Compared to the fiducial model, we see that the model with a lower galaxy stellar mass function (M*−σ; 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 to the difference between L1_m9 and M*−σ, the difference between the fgas−4σ and M*−σ_fgas−4σ 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. 2.7 to the stellar mass variations in Fig. 2.8, we see that M*−σ shows a similar suppression as the fgas−2σ model, both showing a suppression of $\approx$ 15% 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*−σ_fgas−4σ model shows some additional suppression compared to fgas−4σ as the former run seems to fall in between fgas−4σ and fgas−8σ, 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 Jet_fgas−4σ 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% 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 stellar mass functions.

2.4.5 Cosmology variations

In this section, we study the effect of varying the cosmology. The top panel in Fig. 2.9 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

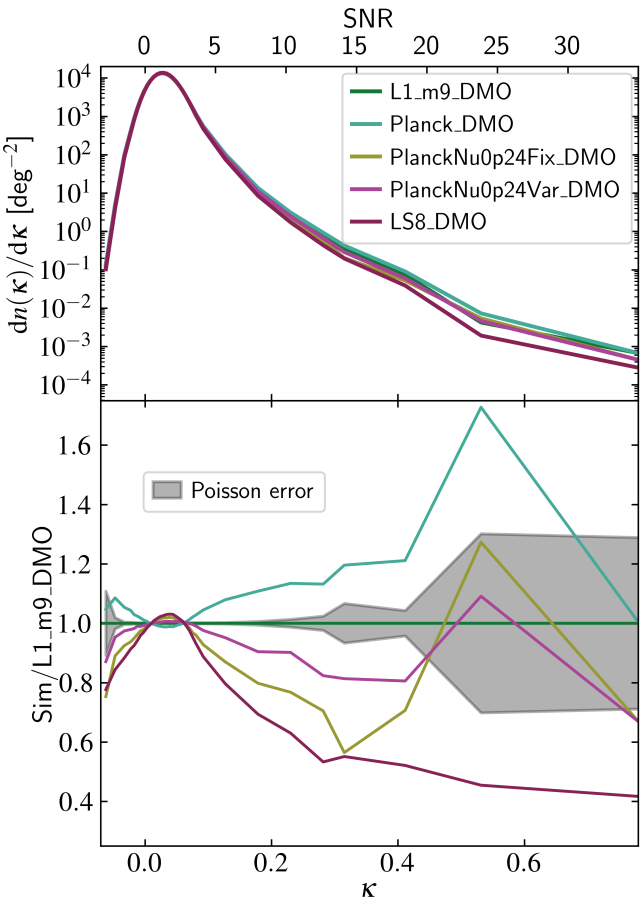


Figure 2.9: As Fig. 2.6 but comparing the cosmology variations in the 1 cGpc DMO box. The cosmologies are listed in Table 2.3. The vertical range in the bottom panel is larger than in all similar plots as the cosmology variations, in general, show larger differences than the baryonic feedback variations. The differences can be qualitatively understood from the impact of cosmological parameters (primarily Ω_m and σ_8) on the HMF.

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., 2024). Compared to the fiducial model, the LS8 model has almost the same value of Ω_m but a 6% lower value of σ_8. We thus expect fewer massive haloes in the LS8 model. This is reflected in Fig. 2.9, 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 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 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 to 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. 2.9, 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%. 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 SNR = 0 to 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% lower compared to the Planck model. In cosmology, we understand neutrinos to act as a form of hot dark matter (HDM). 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 $10 - 20\%$ 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% 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 $\sum m_\nu = 0.06$ to 0.24). They too find a suppression of $20 - 30\%$ 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 $\sum m_\nu = 0.06$ and 0.24 eV. They see a difference of $10 - 20\%$ up to 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 $\approx 20 - 60\%$. The differences can be understood qualitatively by considering the impact of cosmological parameters on the HMF.

2.4.6 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. 2.10, 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. 2.9, when comparing the variations to their own DMO run, the baryonic suppression is very similar for all cosmology variations, as shown in the bottom panel of Fig. 2.10, 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) %, and generally falls within the Poisson noise for L1_m9 as indicated by the gray-shaded area, whereas the cosmology variations in Fig. 2.9 show up to an order of magnitude larger differences in the same κ regime.

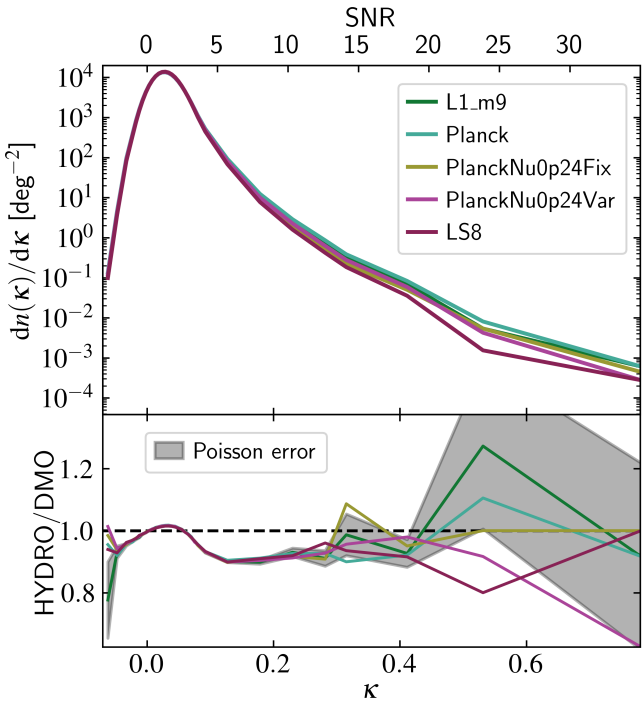


Figure 2.10: As Fig. 2.6 but comparing the hydrodynamical runs with different cosmologies but the same baryonic calibration to their corresponding DMO runs. Whereas the variation between the DMO signals in Fig. 2.9 is up to 300%, the baryonic suppression (as seen in the bottom panel) is similar for all cosmology variations, illustrating the separability of the cosmology and baryonic impact.

We can understand the separability from the different impacts of the two processes on halos. We have seen that the cosmology dependence translates itself into changes in the total number of halos 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 centers of halos. 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.

2.5 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. 2.6 shows that the fiducial baryonic suppression is $\approx 10\%$ for $\kappa = 0.1$ and that the suppression decreases for larger WL convergence values. Comparing this to the suppression and enhancement of the different feedback variations in Fig. 2.7 and 2.8, we see that the other baryonic feedback variations show a suppression of 5-30% compared to 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 modeled 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. 2.9 are larger than the differences between the runs that vary the gas fractions, as seen in Fig. 2.7, which span 10σ in gas fractions. The enhancement in the number density of the Planck cosmology is about twice as large as for the fgas+2σ model and LS8 suppresses the number density of peaks $2-3$ times as much as the fgas-8σ 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 stellar mass function 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 $\Omega_{\rm m}$, and its clumpiness, as set by S_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 to 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 to the fiducial

model (Fig. 2.7) and the heavier neutrino models to the Planck model (Fig. 2.9), 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 fgas+2σ and fgas−8σ. Our analysis thus shows that WL peak counts are more sensitive to the difference in cosmological parameters between the two sets of cosmologies currently favored 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 stellar mass functions.

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 modeled. In particular, the impact of baryons and cosmology shows up in the same convergence regime with similar behavior, but Section 2.4.6 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 2-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. 2.9). 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. 2.6) 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 kappa regime. We aim to explore both the constraining power and the physical origin of these peaks in future research.

2.6 Summary and Conclusions

We have studied the impact of baryonic feedback on the distribution of peaks in weak lensing 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 stellar mass function and the gas fraction in low$-z$ clusters. To mimic the signal of a Stage IV weak lensing 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. 2.4) and cosmic variance (Fig. 2.5), showing that both are under control in the intermediate weak lensing 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 centers of haloes and thereby decrease their mass, lead to a suppression of the number density of WL peaks of ∼10% in the cosmologically informative WL convergence regime (Fig. 2.6).

At fixed cosmology, initial conditions, and observer position, we compared 9 models that differ in their galaxy stellar mass function, cluster gas fraction, and/or AGN feedback implementation; we find baryonic suppressions in the range 5-30% (Fig. 2.7 & 2.8). 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 center, 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 stellar mass function 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, favored 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)% (Fig. 2.9).

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 modeled 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)% (Fig. 2.10) 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 modeled. 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.

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.

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

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, Journal of High Energy Astrophysics, 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, The Open Journal of Astrophysics, 6, 39
Barreira A., Bose S., Li B., Llinares C., 2017, J. Cosmology 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. Cosmology 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., Ploeckinger 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 et al., 2019, A&A, 627, A59
Euclid Collaboration et al., 2020, A&A, 642, A191
Euclid Collaboration et al., 2023, A&A, 675, A120
Fabbian G., Calabrese M., Carbone C., 2018, J. Cosmology 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. Cosmology Astropart. Phys., 2019, 052

Ferlito F., et al., 2023, MNRAS, 524, 5591
Fluri J., Kacprzak T., Sgier R., Refregier A., Amara A., 2018, J. Cosmology 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, MUSIC2-monofonIC: 3LPT initial condition generator, Astrophysics Source Code Library, record ascl:2008.024 (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, Annual Review of Nuclear and Particle Science, 58, 99
Hu W., 2000, Phys. Rev. D, 62, 043007
Huško 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, Reports on Progress in Physics, 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., Huffenberger K., 2012, Phys. Rev. D, 85, 103513
Kugel R., et al., 2023, MNRAS, 526, 6103
Kuijken K., et al., 2019, A&A, 625, A2
LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201
Lahav O., Liddle A. R., 2022, arXiv e-prints, p. arXiv:2201.08666
Laureijs R., et al., 2011, arXiv e-prints, p. arXiv: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. Cosmology Astropart. Phys., 2018, 049
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Lu T., Haiman Z., 2021, MNRAS, 506, 3406
Marques G. A., et al., 2024, MNRAS, 528, 4513
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
McCarthy I. G., Bird S., Schaye J., Harnois-Deraps J., Font A. S., van Waerbeke L., 2018, MNRAS, 476, 2999
McCarthy I. G., et al., 2023, MNRAS,
McClintock T., et al., 2019, ApJ, 872, 53
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 et al., 2020, A&A, 641, A6
Ploeckinger S., Schaye J., 2020, MNRAS, 497, 4857
Pratten G., Lewis A., 2016, J. Cosmology 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., 2024, MNRAS, 530, 2378
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. Cosmology Astropart. Phys., 2019, 020
Schneider A., Stoira N., Refregier A., Weiss A. J., Knabenhans M., Stadel J., Teyssier R., 2020, J. Cosmology 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, arXiv e-prints, p. 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., 2024, MNRAS, 529, 1862
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. Cosmology 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., 2024, MNRAS, 530, 4203
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., Huffenberger 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, Journal of Open Source Software, 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. Cosmology 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

2.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 6 times to cover the distance to $z = 3$, the approximate redshift range necessary for a Stage IV WL survey (see Fig. 2.1 and the green arrow indicating the L1 box length). As the lightcone extends to $z = 3$ in all directions, in total, 13 L1 box replications are necessary for the L1 runs.

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. 2.1. However, as the lightcone 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. 2.1). 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 6 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 lightcone 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. 2.11 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, dot-

ted, and dashed lines, respectively. The bottom panel compares the three rotation strategies to 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. (2024), who studied similar rotation strategies for the CMB lensing angular power spectrum. All three rotations show some minor suppression compared to 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% 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.

2.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 2.15) 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. 2.12. If smoothing is applied, the map is smoothed with a Gaussian kernel with FWHM = 1 arcmin. The figure includes the mean of the 8 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. 2.12 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%. 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 to 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

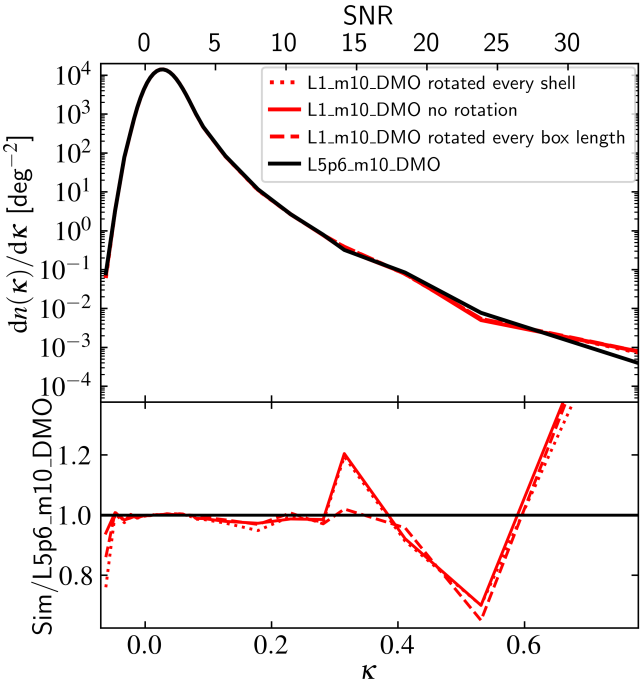


Figure 2.11: Top: number densities of WL peaks for the L5p6_m10_DMO model (black) and L1_m10_DMO run (red) with different mass shell rotation strategies where the shells in the latter simulation are not rotated (dashed), every shell is randomly rotated (dotted) or the shells are randomly rotated every half box length (solid). Bottom: ratio to the L5p6_m10_DMO signal. The different rotation methods only show minor differences.

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 to observations a more dedicated study into these systematics should be carried out.

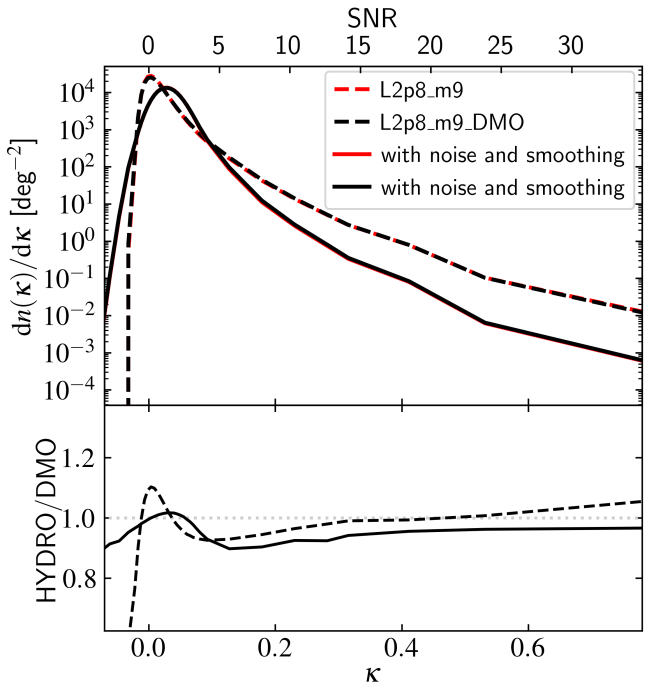


Figure 2.12: Top: number densities of WL peaks for the mean of the 8 observers in the L2p8_m9 (red) and L2p8_m9_DMO (black) runs with different treatment for the applied smoothing and noise. The solid and dashed lines correspond to number densities measured from maps that have no smoothing or noise, 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.

3 The FLAMINGO project: cosmology with the redshift dependence of weak gravitational lensing peaks

Authors: **Jeger C. Broxterman**, Matthieu Schaller, Henk Hoekstra, Joop Schaye, Robert J. McGibbon, Victor J. Forouhar Moreno, Roi Kugel, and Willem Elbers
Published in Monthly Notices of the Royal Astronomical Society, 538, 755 (2025)

Abstract

Weak gravitational lensing (WL) convergence peaks contain valuable cosmological information in the regime of non-linear collapse. Using the FLAMINGO suite of cosmological hydrodynamical simulations, we study the physical origin and redshift distributions of the objects generating WL peaks selected from a WL convergence map mimicking a *Euclid* signal. We match peaks to individual haloes and show that the high signal-to-noise ratio (SNR > 5) WL peaks measured by Stage IV WL surveys primarily trace $M_{200c} > 10^{14}$ $M_\odot$ haloes. We find that the WL peak sample can compete with the purity and completeness of state-of-the-art X-ray and Sunyaev-Zeldovich cluster abundance inferences. By comparing the distributions predicted by simulation variations that have been calibrated to the observed gas fractions of local clusters and the present-day galaxy stellar mass function, or shifted versions of these, we illustrate that the shape of the redshift distribution of SNR > 5 peaks is insensitive to baryonic physics while it does change with cosmology. The difference highlights the potential of using WL peaks to constrain cosmology. As the WL convergence and redshift number densities of WL peaks scale differently with cosmology and baryonic feedback, WL peak statistics can simultaneously calibrate baryonic feedback and constrain cosmology.

3.1 Introduction

The cornerstone of modern cosmology is the ΛCDM model, consisting primarily of dark energy (Λ) and cold dark matter (CDM). The model can predict a wide range of cosmological observables, including the cosmic microwave background (CMB, e.g. Planck Collaboration et al., 2020a), baryonic acoustic oscillations (BAO, e.g. DESI Collaboration et al., 2025), and measurements of the luminosity distance (D_L) from supernovae (SNe, e.g. Riess et al., 2021). Other probes, such as redshift-space distortions (RSDs) (e.g. Scoccimarro, 2004), galaxy clustering (e.g. Cacciato et al., 2009), CMB lensing (e.g. Planck Collaboration et al., 2020b), and cosmic shear measure the properties of the large-scale structure (LSS) over a range of scales and with cosmic time. Cosmic shear is the distortion of galaxy shapes by weak gravitational lensing (WL) by the large-scale structure, and it is a powerful tool to map the matter distribution in the universe (for a review, see e.g. Kilbinger, 2015). Many of the LSS probes report values of the S_8 parameter that are in mild tension with CMB predictions. This parameter is well constrained by WL inferences and measures a combination of the matter density and the amplitude of density fluctuations. The S_8 tension manifests itself between CMB and cosmic shear (e.g. Hildebrandt et al., 2017; Asgari et al., 2021; Secco et al., 2022; Dark Energy Survey and Kilo-Degree Survey Collaboration et al., 2023), galaxy clustering (e.g. Tröster et al., 2020; Philcox & Ivanov, 2022), RSDs (e.g. Benisty, 2021), and thermal Sunyaev-Zeldovich (tSZ) measurements (e.g. McCarthy et al., 2014, 2023).

An additional statistic that is used to constrain cosmology is cluster counts (e.g. Wang & Steinhardt, 1998; Reiprich & Böhringer, 2002; Mana et al., 2013; Fumagalli et al., 2024). Clusters are the largest virialized structures in the universe, and their number density probes the high-mass end of the halo mass function (HMF), which is sensitive to the amount of matter (Ω_m), its clumpiness (σ_8), and the evolution of dark energy (w_0, w_a) (e.g. Courtin et al., 2011; Bocquet et al., 2016). For a review, see Allen et al. (2011).

Clusters may be selected through different observables (e.g. optical richness, SZ or X-ray), but selection effects and model assumptions typically limit cluster cosmology inferences (e.g. Nagai et al., 2007; Kugel et al., 2025). Whereas a Compton-y selected sample can provide cleaner selection effects, and more accurate mass estimates can be obtained by calibrating mass proxies using lensing measurements, the approach still suffers from not being able to definitively determine the completeness of the sample as clusters may be missing from the sample due to variations in their temperature and gas content (e.g. Kugel et al., 2024). Additionally, current cluster inferences require several model assumptions, such as the choice of HMF, the assumed functional form for observable-mass scaling relations and the associated scatter, which may lead to biased results for next-generation measurements that aim to do precision cosmology at the sub-per cent

level (Bocquet et al., 2016; Kugel et al., 2025).

In this paper, we instead explore a different approach that aims to extract cosmological information by also probing the high-mass end of the HMF as a function of redshift, but that suffers less from selection effects and model assumptions as the selection is carried out using a WL convergence (κ) map only. By selecting peaks from a WL map, we expect to extract the same (and more) information probed by current cluster analyses, as we will show that the peaks correspond to virtually all the most massive objects in the universe leading to a pure and complete sample. Crucially, the WL selection of these objects is not plagued by the uncertainties originating from assumptions regarding the dynamical state of the galaxy clusters (e.g. relaxedness or hydrostatic equilibrium) nor by scaling relation calibration problems.

This paper focuses on the local maxima in WL convergence maps, commonly referred to as WL peaks (see e.g. Dietrich & Hartlap, 2010; Kacprzak et al., 2016; Martinet et al., 2018; Davies et al., 2021, 2022; Li et al., 2023). These peaks are a non-Gaussian statistic and correspond to the most massive objects in the universe that have undergone non-linear collapse and, therefore, contain a wealth of cosmological information (see e.g. Kratochvil et al., 2010). The origin of WL peaks in the context of single, massive objects has been studied using simulations by Hamana et al. (2004), who found that high signal-to-noise ratio (SNR) peaks originate from single massive objects along the line of sight as opposed to several haloes aligned along the line of sight. These results were confirmed by Yang et al. (2011) and Liu & Haiman (2016), who also found that peaks with SNR > 3.5 can generally be attributed to a single massive halo. The studies broadly agree, but as there are differences in redshift distributions, noise values, methodology, and SNR cuts, the three studies differ in their exact predictions. We aim to further this understanding by considering the analysis in the context of Stage IV WL surveys, i.e. *Euclid* (Euclid Collaboration: Mellier et al., 2025), *Roman* (Spergel et al., 2015), and *Rubin* (LSST Science Collaboration et al., 2009), and by matching individual peaks to haloes to determine the halo mass regimes corresponding to WL peaks of different WL convergence values.

The WL convergence value of a WL peak depends on the lens mass and redshift (as this impacts the WL kernel) and consequently depends on structure formation as more massive structures form at lower redshift, but also on baryonic feedback as this removes mass from the haloes that act as lenses (see e.g. Stanek et al., 2009; Martizzi et al., 2014; Velliscig et al., 2014; Debackere et al., 2022). In Broxterman et al. (2024), we studied how baryonic physics and changes in cosmology impact the counts of WL peaks using the FLAMINGO simulation suite (Schaye et al., 2023; Kugel et al., 2023). There, we showed that the impact of baryonic physics and cosmology on the number density of high-valued WL peaks can be qualitatively understood by considering the impact of feedback processes on the mass of a single halo and the impact of the cosmological parameters (primar-

ily Ω_m and σ_8) on the HMF. Feedback removes gas from the centres of haloes, causing them to be less overdense. Consequently, the WL effect will be smaller, and the peaks they generate will have lower WL convergence values. Hence, there will be fewer (high-valued) κ peaks when haloes experience stronger feedback. Regarding the impact of cosmology, Ω_m broadly moves the amplitude of the HMF vertically, and σ_8 shifts its exponential cut-off, both causing there to be more/fewer massive haloes in the universe in general, which in turn is also reflected in the number of (high-valued) κ peaks. Whereas the impact of realistic baryonic physics is smaller than currently reasonable changes in cosmology, the two signatures are degenerate, complicating a hypothetical cosmological inference.

In this work, we aim to extend this analysis by including the redshift information of the WL peaks. While the number density of WL peaks in κ depends on cosmology and baryonic physics, we expect the formation time of massive objects and corresponding WL peak redshift to depend primarily on cosmology. The formation time of massive haloes should, to first order, not depend on the exact strength of baryonic feedback. Therefore, the redshift distribution of the peaks contains information on when and how much structure in the universe has formed that could be used as an additional probe to constrain cosmology. The redshift dependence of the number density of WL peaks corresponding to these massive objects should be insensitive to baryonic physics. Therefore, we aim to explore the possibility of using the redshift dependence of the number density of WL peaks to constrain cosmology and discriminate between the impact of baryonic feedback and changes in cosmology.

Recently, Harnois-Déraps et al. (2024) carried out an inference using WL peaks selected on tomographic WL convergence maps from the KiDS-1000 and DES Y1 data releases. They found that this can greatly improve the results compared with selecting peaks from a single integrated WL convergence map as it can probe the evolution of the growth of structure, already highlighting the possibility of including redshift information. We aim to assign a redshift to every peak and thus expect to include more redshift information than in their study. Also, Chen et al. (2025) and Chiu et al. (2024) carried out a similar analysis as the idea explored in this paper by using HSC Y3 data and by selecting clusters through the peaks of WL aperture mass maps. Their sample only consists of 149 clusters selected through WL, and they thus cannot yet provide meaningful constraints. Nonetheless, it highlights the viability of the approach. For a recent review of peaks selected from aperture mass maps, see Oguri & Miyazaki (2025).

This paper aims to explore in more detail the origin of WL peaks by linking them to individual haloes and leveraging the baryonic insensitivity of their redshift distribution to illustrate their potential to better constrain cosmological parameters. To this end, we use the baryonic physics and cosmology variations of the cosmological hydrodynamical simulation suite FLAMINGO (Schaye et al.,

2023; Kugel et al., 2023). We first briefly summarize the relevant WL theory in Section 3.2. We generate full-sky WL convergence maps mimicking the signal of Stage IV WL surveys, as described in Section 3.3, where we also introduce the algorithm to match individual peaks to massive haloes each. The results of the halo-peak matching and the evolution of the redshift number density of WL peaks are presented in Section 3.4. The main conclusions are summarized in Section 3.5.

3.2 Theory

As we will construct full-sky WL maps from discretized mass shells spaced linearly in redshift, we first summarize the main equations relevant to WL. For full reviews of the WL formalism, see Bartelmann & Schneider (2001); Hoekstra & Jain (2008); Kilbinger (2015). In WL, where the light deflections are small, the magnification matrix ($\boldsymbol{A}$), which describes the linear transformation from the observed angular position ($\boldsymbol{\theta}$) to the unlensed position, is typically decomposed as

$$A_{ij} = \frac{\partial(\delta\theta_i)}{\delta\theta_j} \approx \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{3.1}$$

where κ is the WL convergence and $\gamma = \gamma_1 + i\gamma_2$ is the WL shear. When assuming the deflections are small, the deflection angle can be expressed as the gradient of the potential (Φ) that is the solution of the 3D Poisson equation, $\nabla^2\Phi = 4\pi G a^2 \bar{\rho_\mathrm{m}}\delta$, where G is the gravitational constant, a the scale factor, $\bar{\rho}$ the mean density of the universe, and δ the matter overdensity. By assuming the Born approximation, i.e. evaluating the lensing equations on unperturbed photon paths, the magnification matrix reduces to

$$A_{ij} = \delta_{ij} - \partial_i\partial_j\psi, \tag{3.2}$$

with ψ the lensing potential,

$$\psi(\boldsymbol{\theta},\chi) = \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \, \frac{\chi-\chi'}{\chi\chi'}\,\Phi(\chi'\boldsymbol{\theta},\chi'), \tag{3.3}$$

where c is the speed of light, and we assumed a flat universe such that the comoving angular diameter distance, $f_K(\chi)$, reduces to the comoving distance, χ. By substituting the Poisson equation, the WL convergence can be estimated as

$$\kappa(\boldsymbol{\theta},\chi) = \frac{3\Omega_\mathrm{m}H_0^2}{2c^2}\int_0^{\chi_\mathrm{hor}} \mathrm{d}\chi\,[1+z(\chi)]\,W(\chi)\,\delta(\chi,\boldsymbol{\theta}), \tag{3.4}$$

where H_0 is the Hubble constant, $\Omega_\mathrm{m} \equiv \rho_{\mathrm{m},0}/\rho_{\mathrm{crit},0}$ is the matter density parameter, $\rho_{\mathrm{crit},0} = 3H_0^2/(8\pi G)$ is the present-day critical density of the universe, and z is

the redshift. $\chi_{\rm hor}$ is the comoving line of sight horizon to the edge of the galaxy sample. The weak lensing kernel W is given by (Kaiser, 1992)

$$W(\chi) = \chi \int_{\chi}^{\chi_{\rm hor}} \mathrm{d}\chi' \, n_{\rm s}(\chi') \frac{\chi' - \chi}{\chi'}, \tag{3.5}$$

where $n_{\rm s}(\chi)$ is the comoving source distribution which is normalised ($\int n_{\rm s}(\chi)\mathrm{d}\chi = 1$) and is related to the source redshift distribution as $n_{\rm s}(z)\mathrm{d}z = n_{\rm s}(\chi)\mathrm{d}\chi$.

Lu & Haiman (2021) have argued that for a Stage III WL peak inference assuming the Born approximation leads to biased results, and a more sophisticated approach such as backward ray tracing should be applied. In this paper, we do not aim to carry out parameter inference but are instead interested in the origin and redshift evolution of a sample of high-valued WL peaks. Although the exact number of WL peaks at fixed WL convergence values may differ slightly depending on the map construction algorithm, we do not expect the origin of these peaks to change, as they should still stem from the same massive structures in the universe. As the Born approximation facilitates a more straightforward and computationally less expensive way to assign a redshift (as will become clear in Section 3.3.2), we adopt it in our analysis. Additionally, as we ignore several second-order effects such as lens-lens coupling (Bernardeau et al., 1997), and possible biasing factors such as dilution by cluster member contamination (Applegate et al., 2014; Medezinski et al., 2018), the analysis is to some extent idealized. Some of these effects, such as source clustering, are expected to have a minor impact on WL peak statistics (Gatti et al., 2023). For others, such as galaxy intrinsic alignment, the impact, although small, is non-negligible (Zhang et al., 2022). As the impact of these effects is less than 1 per cent on the power spectrum (Kilbinger, 2015), we do not expect them to impact the conclusions of this work as we study the relative differences between the different simulation variations, but an analysis using legacy Stage IV WL survey data should consider them in more detail.

3.3 Methods

3.3.1 FLAMINGO

We use the recently developed FLAMINGO cosmological hydrodynamical simulation suite for our analysis. In this section, we only provide a brief summary of the simulation characteristics; for a detailed description of the simulations and calibration, see Schaye et al. (2023) and Kugel et al. (2023). The simulations were run using the SPHENIX smoothed particle hydrodynamics implementation (Borrow et al., 2022) in SWIFT (Schaller et al., 2024). The simulations include an element-by-element based radiative cooling and heating (Ploeckinger & Schaye, 2020), star formation (Schaye & Dalla Vecchia, 2008), time-dependent stellar

mass loss (Wiersma et al., 2009), and massive neutrinos (Elbers et al., 2021). Supernova and stellar feedback is kinetic and conserves energy and linear and angular momentum (Dalla Vecchia & Schaye, 2008; Chaikin et al., 2023). The AGN feedback is either thermal (Booth & Schaye, 2009) or kinetic jet-like, where particles are kicked along the black hole (BH) spin axis (Huško et al., 2022). The simulations are calibrated using machine learning to the galaxy stellar mass function at $z = 0$ and the gas fraction of low-z clusters and groups of clusters, or systematic shifts in these observables, while taking into account observational uncertainties and biases (Kugel et al., 2023).

In Table 3.1, we list the simulation parameters directly relevant to the interpretation of our results. The first column lists the simulation's identifier, which is set by the box size in comoving Gpc (cGpc) and $\log_{10}$ baryonic particle mass. For example, the flagship simulation, L2p8_m9, has 5040^3 baryonic and dark matter (DM) particles and 2800^3 massive neutrino particles in a 2.8 cGpc box and an initial mean gas particle mass of 1.07×10^9 $M_\odot$. All 7 cosmology and 8 baryonic feedback variations were carried out in 1 cGpc boxes with an equal number of particles. The fiducial model in this box is L1_m9. L1_m8 and L1_m10 are variations with the same cosmology and calibrated to the same observational data but with 8 times more and fewer particles, respectively. We will use these to quantify the numerical convergence with resolution. All cosmology variations have an accompanying CDM + neutrino run identified by the post-fix _DMO.

The eight variations with different baryonic feedback prescriptions were each calibrated to a (shifted) set of observables, as indicated by the fourth and fifth columns in Table 3.1. These variations vary the cluster gas fractions ($\Delta f_{\rm gas}$) and/or the galaxy stellar mass function (ΔM_*) by a set amount of observational standard deviations (σ). The baryonic variations are identified by their shifts in observables, as indicated in the first column of Table 3.1.

All baryonic variations assume the DES Y3 3 × 2pt + All Ext. best-fitting ΛCDM cosmology from Abbott et al. (2022), see also Table 3.1. Additional cosmology variations were run using the Planck Collaboration et al. (2020a) best-fitting ΛCDM model assuming $\sum m_\nu c^2 = 0.06$ eV (Planck), and variations with heavier neutrinos, $\sum m_\nu c^2 = 0.24$ eV ('PlanckNu0p24Var', 'PlanckNu0p24Fix') and $\sum m_\nu c^2 = 0.48$ eV ('PlanckNu0p48Fix'). The PlanckNu0p24Var and PlanckNu0p24Fix differ in their treatment of the other cosmological parameters while changing the neutrino mass. We include two variations with decaying cold dark matter (DCDM) from Elbers et al. (2025). In these simulations, dark matter decays to dark radiation as specified by the decay rate Γ in units H_0/h, as indicated by the final column in Table 3.1. Finally, we use a simulation with a lensing cosmology with a lower value of σ_8 ('LS8') from Amon et al. (2023) that is consistent with BOSS DR12 galaxy clustering (Reid et al., 2016), and galaxy-galaxy lensing from KiDS-1000 (Asgari et al., 2021), DES Y3 (Abbott et al., 2022), and HSC Y1 (Aihara et al., 2018).

Table 3.1: Simulation variations. From left to right, the columns list the simulation identifier; the box size, L; the number of baryonic and dark matter particles, N_p; the initial mean baryonic particle mass, m_b; the shift in the number of standard deviations (σ) applied to the observational data used in the calibration for the $z = 0$ galaxy stellar mass function (ΔM_*) and the gas fraction of low-redshift clusters (Δf_gas); the mode of AGN feedback; the dimensionless Hubble constant, h; the present-day total matter density parameter, Ω_m; the present-day baryonic matter density parameter, Ω_b; the sum of neutrino species particle masses, $\sum m_\nu c^2$; the amplitude of the initial power spectrum parametrized as the rms mass density fluctuation in spheres with radius 8 Mpc/h extrapolated to $z = 0$ using linear theory, σ_8 and, if applicable, the dark matter decay rate, Γ.

Identifier	L	N_p	m_b	ΔM_*	Δf_gas	AGN	h	Ω_m	Ω_b	$\sum m_\nu c^2$	σ_8	Γ
	cGpc		$\times 10^9$ M$_\odot$	σ	σ					eV		H_0/h
L2p8_m9	2.8	5040^3	1.07	0	0	thermal	0.681	0.306	0.0486	0.06	0.807	–
L1_m9	1	1800^3	1.07	0	0	thermal	0.681	0.306	0.0486	0.06	0.807	–
L1_m8	1	3600^3	0.134	0	0	thermal	0.681	0.306	0.0486	0.06	0.807	–
L1_m10	1	900^3	8.56	0	0	thermal	0.681	0.306	0.0486	0.06	0.807	–
fgas+2σ	1	1800^3	1.07	0	+2	thermal	0.681	0.306	0.0486	0.06	0.807	–
fgas−2σ	1	1800^3	1.07	0	−2	thermal	0.681	0.306	0.0486	0.06	0.807	–
fgas−4σ	1	1800^3	1.07	0	−4	thermal	0.681	0.306	0.0486	0.06	0.807	–
fgas−8σ	1	1800^3	1.07	0	−8	thermal	0.681	0.306	0.0486	0.06	0.807	–
M*−σ	1	1800^3	1.07	−1	0	thermal	0.681	0.306	0.0486	0.06	0.807	–
M*−σ_fgas−4σ	1	1800^3	1.07	−1	−4	thermal	0.681	0.306	0.0486	0.06	0.807	–
Jet	1	1800^3	1.07	0	0	jet	0.681	0.306	0.0486	0.06	0.807	–
Jet_fgas−4σ	1	1800^3	1.07	0	−4	jet	0.681	0.306	0.0486	0.06	0.807	–
Planck	1	1800^3	1.07	0	0	thermal	0.673	0.316	0.0494	0.06	0.812	–
PlanckNu0p24Var	1	1800^3	1.06	0	0	thermal	0.662	0.328	0.0510	0.24	0.772	–
PlanckNu0p24Fix	1	1800^3	1.07	0	0	thermal	0.673	0.316	0.0494	0.24	0.769	–
PlanckNu0p48Fix	1	1800^3	1.07	0	0	thermal	0.673	0.316	0.0494	0.48	0.709	–
PlanckDCDM12	1	1800^3	1.07	0	0	thermal	0.673	0.274	0.0494	0.06	0.794	0.12
PlanckDCDM24	1	1800^3	1.07	0	0	thermal	0.673	0.239	0.0494	0.06	0.777	0.24
LS8	1	1800^3	1.07	0	0	thermal	0.682	0.305	0.0473	0.06	0.760	–

Mass maps

We use the FLAMINGO mass maps to construct our WL convergence maps. These are a set of concentric spherical HEALPix (Górski et al., 2005) surfaces on which the particles that contribute to an observer's past light-cone are projected (see Appendix A2 of Schaye et al., 2023). The L1 and L2p8 simulations have 2 and 8 virtual observers, respectively. The light-cone consists of 60 shells of equal thickness (in redshift) between $z = 0$ and 3 ($\Delta z = 0.05$). These maps are the input for constructing the WL convergence maps. To allow the use of all functionality of HEALPY (Zonca et al., 2019), we downsample the FLAMINGO maps to $N_{\rm side} = 8192$, corresponding to an angular resolution of ≈ 25.8 arcsec. We then transform the total matter shells at each shell i into overdensity shells, $\delta_i(\boldsymbol{\theta})$, using

$$\delta_i(\boldsymbol{\theta}) = \frac{\Sigma_i(\boldsymbol{\theta}) - \overline{\Sigma_i}}{\overline{\Sigma_i}}, \tag{3.6}$$

where $\Sigma_i(\boldsymbol{\theta})$ is the surface matter density at shell i at position $\boldsymbol{\theta}$ and $\overline{\Sigma_i}$ is the mean surface matter density, which we evaluate directly from the shell. Because of the discrete redshifts of the FLAMINGO mass maps, we transform the analytic expressions to estimate the WL convergence (equations 3.4 & 3.5) to discrete sums given by (e.g. McCarthy et al., 2018)

$$\kappa(\boldsymbol{\theta}) = \frac{3\Omega_{\rm m} H_0^2}{2c^2} \sum_{i=1}^{N} [1 + z(\chi_i)]\, W(\chi_i)\, \delta_i(\boldsymbol{\theta})\, \Delta\chi_i, \tag{3.7}$$

and

$$W(\chi_i) = \chi_i \sum_{j=i}^{N} \Delta\chi_j\, n_{\rm s}(\chi_j)\, \frac{\chi_j - \chi}{\chi_j}, \tag{3.8}$$

respectively. In our analysis, we assume a *Euclid*-like source redshift distribution given by (Euclid Collaboration: Blanchard et al., 2020):

$$n_{\rm s}(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right], \tag{3.9}$$

with $z_0 = 0.9/\sqrt{2}$.

We randomly rotate the mass maps whenever the light-cone diameter exceeds the box size to avoid encountering the same structure multiple times along the line of sight but, at the same time, not unnecessarily erasing correlations along the line of sight. In Appendix A of Broxterman et al. (2024) we found that the rotation prescription had little impact on the final number density of WL peaks. To validate our map construction procedure, we first compare the angular power

spectrum of the final WL convergence map to the prediction from HALOFIT (Takahashi et al., 2020) provided by CLASS (Blas et al., 2011) in Fig. 3.1. CLASS provides the three-dimensional matter power spectrum ($P_{\rm m}$), which we relate to the angular power spectrum ($\mathcal{C}(\ell)$) using (Limber, 1953; LoVerde & Afshordi, 2008):

$$\mathcal{C}(\ell) = \frac{9H_0^4\Omega_{\rm m}^2}{4c^4}\int_0^{\chi_{\rm hor}} {\rm d}\chi\, \frac{W^2(\chi)}{\chi^2 a^2(\chi)}\, P_{\rm m}\left(\frac{\ell+1/2}{\chi}, z(\chi)\right), \tag{3.10}$$

where $W(\chi)$ is given by equation 3.5. The HALOFIT + CLASS prediction is given by the black dashed curve. The green solid line shows the estimate using the employed methods in this paper, i.e. assuming the Born approximation and using the HEALPY ANAFAST routine. Finally, the dotted gray curve shows the angular power spectrum generated using the more elaborate backward ray-tracing method used in Broxterman et al. (2024). The bottom panel shows the ratio to the CLASS + HALOFIT prediction. At the power spectrum level, the ray tracing and Born approximation give the same results, as expected (Hilbert et al., 2020). The estimates also agree well with the HALOFIT prediction and illustrate that the map construction algorithm works well. The difference at small scales is caused by the pixelization of the matter field onto the HEALPix grid, as discussed in more detail in Section 3.5 of Broxterman et al. (2024). For a detailed comparison between predictions from HALOFIT and FLAMINGO, see Upadhye et al. (2024).

3.3.2 Weak-lensing peaks

To mimic a *Euclid*-like signal, we smooth the final maps with a Gaussian beam with a smoothing scale of 1 arcmin and we apply galaxy shape noise by drawing from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = \sigma_\epsilon/\sqrt{2n_{\rm gal}A_{\rm pix}}$, where $\sigma_\epsilon = 0.26$ is the rms total intrinsic ellipticity of the source galaxies, $n_{\rm gal} = 30$ arcmin^{-2} is the source galaxy number density and $A_{\rm pix}$ is the area of a HEALpix pixel (Kaiser & Squires, 1993; van Waerbeke, 2000; Euclid Collaboration: Martinet et al., 2019; Euclid Collaboration: Ajani et al., 2023; Euclid Collaboration: Mellier et al., 2025).[1] By smoothing the maps with different smoothing lengths, information corresponding to different scales may be probed, as illustrated by Li et al. (2019). In this work, we focus on a single smoothing scale but note that the main conclusions of the halo-peak matching are independent of the exact value of the smoothing scale, as illustrated in Appendix 3.A. We define a peak as any pixel on the HEALPix grid with a value greater than those of its 8 nearest neighbors. Although a crude definition, this has been used widely in observations (e.g. Hamana et al., 2015; Shan et al., 2018; Grandón et al., 2024)

[1] This noise configuration likely slightly underestimates the actual noise that will be measured by Stage IV WL surveys but we still adopt it to facilitate a consistent comparison with Broxterman et al. (2024).

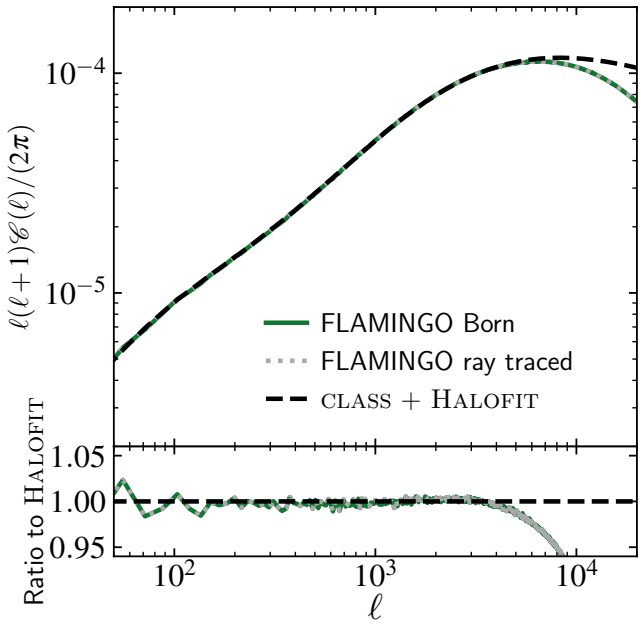


Figure 3.1: Top: Weak lensing convergence angular power spectrum estimated using a *Euclid*-like source redshift distribution, corresponding to the mean of the 8 L2p8_m9_DMO observers. The full-sky map is constructed from the FLAMINGO mass maps and assumes the Born approximation (solid green) or a more advanced backward ray-tracing method (dotted gray). The dashed black curve shows the prediction from HALOFIT using the non-linear 3D matter power spectrum provided by CLASS. Bottom: ratio to the CLASS + HALOFIT prediction. The three estimates agree to an excellent degree apart from small scales due to the pixelisation of the HEALpix maps.

as well as simulations (e.g. Weiss et al., 2019; Coulton et al., 2020; Davies et al., 2024).

To match the peaks to haloes, we assign two angular coordinates and a redshift to each peak and halo (see also Section 3.3.3). The positions of the peaks on the HEALPix grid directly provide the angular coordinates of the peaks. To assign a redshift to each peak, we exploit the summation properties of the Born approximation. As explained in Section 3.3.1, the WL signal is estimated with a double summation (equations 3.7 & 3.8). The first summation sums over all the lens planes, and for each lens plane, the second summation calculates the contribution of all more distant source planes. For each peak, we determine the redshift of the shell where the contribution from the lens (equation 3.7) is the largest and assign this redshift to the peak. Before determining the maxima, we

smooth the contributions at each shell with the same smoothing kernel as applied to the final map. We refer to the redshift assigned in this way as $z_{\rm true}$ as this is the redshift of the major overdensity that acts as the lens to generate the peak. After matching the peaks to individual haloes, as explained in Section 3.3.4, the peaks are assigned the redshift of the halo to which they have been matched. In Section 3.3.5, we will modify these redshifts to account for realistic observational redshift uncertainties.

The straightforwardness with which we can directly assign a redshift to each peak is the main reason why we choose to adopt the Born approximation. Assigning a redshift to a peak is less straightforward using the ray-tracing methodology used in Broxterman et al. (2024), as that approach effectively sums over the contributions of all the sources, instead of the lenses.

Purity

To ensure that we select a peak sample that predominantly corresponds to a single halo along the line of sight towards each peak, we apply cuts based on the peak value, applied noise, and most and second most contributing lens shells, all measured at the angular position of the peak. First, if the smoothed random noise alone can account for over 50 per cent of the final peak value, we call this peak a 'Noise peak'. Second, to guarantee that the object corresponds to a single shell and it is not a result of a superposition of multiple objects along the line of sight, we require the most contributing shell to contribute more than 50 per cent of the final peak value and the second most contributing shell to contribute less than 50 per cent of the most contributing shell. We refer to the subset of peaks satisfying these criteria as 'Single shell peaks'. These cut will throw away a small subset of haloes located directly on the edge of two subsequent shells, which are, in reality, single haloes. As we aim to get a pure selection of peaks, we do not study these peaks in more detail, but we note that, as a consequence, our resulting selection of Single shell peaks will likely be slightly underestimated. We refer to the remaining set of peaks as 'Multiple shells peaks', as their WL convergence cannot be attributed to the applied noise or a single shell, or there are multiple contributing shells along the line of sight.

In Fig. 3.2, we show the fractional contribution of each of the three peak categories to the total number of peaks as a function of their WL peak convergence value. The top panel corresponds to a *Euclid*-like analysis, whereas the bottom panel corresponds to a Stage III WL survey-like analysis. The top axis of each panel shows the SNR ($= \kappa/\sigma_{\rm noise}$). The dashed-dotted, dashed, and solid curves indicate the Noise, Multiple shells, and Single shell peaks, respectively. The curves correspond to the mean estimates of the eight observers in L2p8_m9, but we find the same conclusions apply for the measurement in the DMO runs, i.e. baryonic effects do not change this picture.

Focusing first on the top panel, the highest-valued peaks nearly always belong to the Single shell category, as they are caused by a single overdensity along the line of sight. At $\kappa < 0.4$ (0.2), the Multiple shells (Noise) peaks become increasingly important. Below SNR = 5 ($\kappa < 0.11$), the Noise peaks dominate the signal. This is expected as the noise value necessary to account for over half of the signal is increasingly less likely to be drawn for higher-valued peaks. For $\kappa \lesssim 0$ the Noise peak contribution drops again. In Appendix 3.B, we show in more detail that these peaks do not arise from the noise. We find that separating these negative-valued peaks into Single shell and Multiple shells peaks depends strongly on the exact choice of division criterion, which in this case is the ratio of the second to the most contributing shell and the ratio of the most contributing shell to the final peak value. However, this is only the case for the low-κ peaks, and the $\kappa \gtrsim 0.1$ subset, which contains the WL peaks considered in the remainder of this work, is robustly split into the different categories. The contribution of Multiple shells peaks is ~ 10 per cent at SNR = 5 and decreases monotonically until it vanishes at SNR = 18 ($\kappa = 0.40$).

These results agree with those from Hamana et al. (2004) & Yang et al. (2011) who used a different approach but also found that high-signal peaks (in their case SNR > 3.5) principally stem from a single massive halo each. In Fig. 3.2, we see that their convergence cuts are in the regime where the purity drops steeply as the noise and superposition of multiple objects along the line of sight generate a larger fraction of the peaks for lower κ values. At this point, we have not yet aimed to match individual peaks to haloes, but based on the results of Hamana et al. (2004); Yang et al. (2011); Liu & Haiman (2016), we expect that we should be able to match most of the peaks to a single massive halo each, as we will indeed show in Section 3.3.4.

To illustrate the major improvement of the Stage IV WL surveys, we compare the *Euclid* predictions to those resulting from using the same methods but assuming Stage III characteristics. The red curves in the bottom panel of Fig. 3.2 illustrate the same 3 peak categories, but are based on a WL map constructed using the KiDS DR4 (KiDS-1000) source redshift distribution and noise properties (Kuijken et al., 2019; Asgari et al., 2021). The KiDS signal is dominated by Noise peaks to larger κ values, and therefore, fewer peaks can be confidently assigned to a single halo. The main difference rests in the source galaxy background number density, which is a factor ≈ 5 greater for *Euclid*, and therefore, the mean noise is a factor ≈ 2.2 times greater for the KiDS analysis. Additionally, Stage IV surveys detect more distant galaxies, so the WL signals will also be stronger, while the area of the *Euclid* survey will be $\sim$10 times greater, which impacts the total number of measured objects.

Selecting a higher WL convergence cut gives a purer sample as the fraction of Single shell peaks increases. For an SNR-cut of 5 (8), we find a purity percentage (percentage of Single shell peaks) of 76 (96) per cent for *Euclid* and 56 (100) per

cent for KiDS, assuming a KiDS-like noise such that SNR = 5 corresponds to $\kappa = 0.25$ for the latter predictions. We stress that these percentages are likely underestimated due to the conservative cuts we have applied to assign a peak to a category. Not only the purity of the *Euclid* sample is higher at a fixed WL convergence cut, the total number of peaks with SNR > 5 (8) increases from 140 (4) in the full KiDS footprint ($\approx 1,500$ deg^2) to 5.5×10^4 (6.8×10^3) in the full *Euclid* footprint ($\approx 14,000$ deg^2), illustrating the major improvement of Stage IV surveys as they will measure over two orders of magnitude more high-valued peaks. The KiDS estimate is roughly consistent with the recent HSC Y3 analysis results from Chen et al. (2025) and Chiu et al. (2024), who find 129 SNR > 4.7 peaks on the WL aperture mass maps. While their area is smaller than the area we assume for the KiDS-1000 analysis, their survey is deeper, so we expect that their number roughly agrees with our simple KiDS estimate.

We compare our purity prediction to those from Yang et al. (2011). They report a purity of only 20 per cent for SNR > 3.5 peaks. This is 4 times lower than our purity, but lowering the SNR threshold greatly impacts the purity. Assuming their SNR cut, we still find a purity of 38 per cent. Their method attributes a peak to a halo by assuming NFW profiles for each simulated halo and calculating the expected convergence signal using this profile. They then match peaks and haloes using a maximum angular separation of 1.8 arcmin. If the estimated WL convergence is more than half the final WL peak value, they attribute the peak to a single massive halo. However, they already note that this approach's estimated WL convergence has a fractional bias of 7 per cent and underestimates the WL convergence values. As we do not make any assumptions about the halo profile and the division criteria differ, some differences between the two approaches are expected. In conclusion, we find similar purity percentages to those in previous work.

Hamana et al. (2004) estimate that the mean number of haloes that cause a WL peak with SNR > 4 is 37 per deg^2. At the same SNR cut, we find a lower peak number density of only 11 peaks per square degree. Based on the source redshift distribution, we expect their signal to be stronger than we measure. With their SNR cut, they report a contamination rate, i.e. a percentage of peaks that cannot be attributed to a single halo, of 44 per cent. The main difference between their analysis and ours is the choice of SNR cut. From Fig. 3.2 it is clear that the noise and superposition contributions vanish quickly around this SNR regime. If we apply their SNR cut, we find a similar contamination rate of 47 per cent.

In the remainder of our analysis, we apply an SNR cut to our peak sample, as indicated by the dotted black curve in the top panel of Fig. 3.2. We will only use peaks with an SNR value greater than the cutoff value of SNR = 5 to determine the redshift evolution and perform the halo-peak matching. We choose this value because at larger SNR, Single shell peaks dominate and we thus have a reasonably pure sample. Additionally, we stress that until now, we have been conservative in

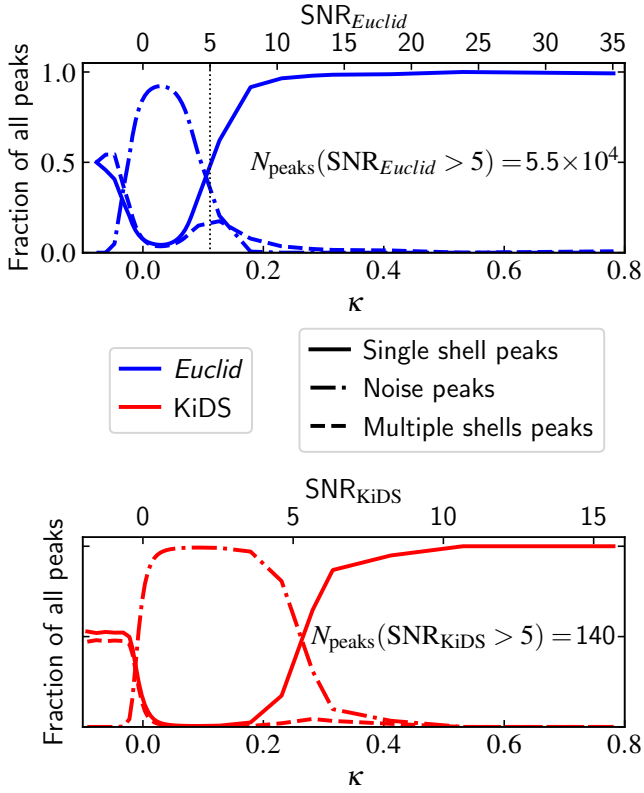


Figure 3.2: Fractional contribution of WL peaks that stem from a single redshift shell ('Single shell', solid), were assigned a random noise that can account for over half of the final peak signal ('Noise peaks', dashed-dotted), or originate from multiple shells along the line of sight ('Multiple shells peaks', dashed) as a function of WL convergence (κ). The blue and red curves correspond to a *Euclid*- and KiDS-like analysis, respectively. The purity (fraction of Single shell peaks) increases with κ. We apply an SNR-cut of 5, as indicated by the vertical dotted line in the top panel, as selection criteria for the remaining analysis. The number of SNR > 5 peaks estimated for the final data releases of the surveys is indicated in each panel, assuming an area of 1,500 deg^2 and 14,000 deg^2, for KiDS and *Euclid*, respectively. Stage IV WL surveys will resolve a large amount of high-SNR peaks corresponding to massive clusters that could be used to constrain cosmology.

our noise and multiple shells cuts as we find that most SNR > 5 peaks belonging to either of these categories still clearly correspond to a single object in the overdensity field, as illustrated in more detail in Appendix 3.C. When studying the redshift distributions of the peaks, we assign a redshift to all Noise and Multiple shells peaks above the SNR-cut and include them in the analysis, as in an observational campaign there is no way of deciding which peaks are generated by the noise.

In Broxterman et al. (2024) we argued that, for a similar analysis, using the number density of WL peaks as a function of their WL convergence strength, the $\kappa = [0.1,0.4]$ regime is most useful for WL peak inferences, as the differences are most easily interpretable in that range. Lower-valued peaks require accurate modeling of instrument and survey characteristics, while the number of higher-valued peaks is subject to significant cosmic variance. Additionally, for this intermediate regime, the differences between the cosmology and baryonic variations could be easily understood by considering the impact of baryonic physics on single haloes and the impact of different cosmological parameters on the HMF. The peaks belonging to this WL convergence regime are predominantly included in our sample after the SNR > 5 cut.

3.3.3 Halo light-cone

To match the peaks to haloes, we use the FLAMINGO halo light-cones. FLAMINGO uses the halo finder HBT-HERONS to identify haloes and track their formation and evolution, including mergers (Han et al., 2018; Forouhar Moreno et al., 2025). The halo finder is used to identify central haloes and their most-bound BH particle, which is set as the centre of the halo. The 3-dimensional positions of the haloes are then converted to angular positions and redshifts, as seen by the different virtual observers in the runs, and saved as halo light-cone files. We directly use these files for our halo-peak matching. We rotate the angular coordinates of the haloes using the same rotations as the light-cone mass maps. On top of the halo finder, which determines the centres of the haloes in the simulation, the FLAMINGO data includes a large set of precomputed quantities in Spherical Overdensity APerture (SOAP) catalogues (McGibbon et al. in prep). From the SOAP catalogues, for each halo, we use M_{200c}, the mass within a spherical aperture with radius R_{200c} enclosing an average overdensity of $\langle\rho(z)\rangle = 200\rho_{\rm crit}(z)$, of the closest snapshot in redshift.

3.3.4 Halo–peak matching

Using the angular coordinates and redshifts of the peaks and haloes, we now match individual peaks to haloes from the FLAMINGO halo light-cones. We per-

form the matching separately for each shell.[2] For each shell, we select all peaks corresponding to that shell. As these peaks all have the same redshift (i.e. the central redshift of the shell), their WL kernel (equation 3.8) is unchanged, and a higher peak should directly originate from a more overdense structure. However, as there may be small variations due to other minor over- or underdense regions within the shell or along the line of sight, the noise that is applied in the end, and because the haloes are not perfectly spherical NFW haloes but differ in shape (e.g. Velliscig et al., 2015; Chua et al., 2019), these peaks will not all correspond to the most massive haloes. Therefore, for each shell, we select the N most massive haloes such that $N_{\rm halo} = 100 N_{\rm peak}$, which are then allowed to be matched to the peaks. We find that changing this value by a factor of two does not impact the results.

To match the haloes to the peaks, we determine the great circle distance ($\Delta\theta$) between each peak and all nearby haloes. First, we project the haloes onto the HEALPix pixels according to their location on the halo light-cone. Then, using the HEALPY function QUERY_DISC, we select all haloes within 2 degrees of the peak. We then calculate the distance between the peak and all haloes in the angular aperture and scale the distances by the quantity $\theta_{\rm 200c} = R_{\rm 200c}/d_{\rm A}$, which is a proxy for the angular size of the halo on the sky, where $d_{\rm A}$ is the angular diameter distance. Then, we select the closest halo in $\Delta\theta/\theta_{\rm 200c}$ for each peak.[3] We do not require a bijective match between the halo and peak as we find that a single halo may cause multiple peaks in the final map, particularly for massive, low-redshift, or merging systems. If, instead, we match peaks to the closest halo in $\Delta\theta/\theta_{\rm 200c}$ without first assigning the peak to a redshift, i.e. matching to all haloes in the same angular region we find that 95 (97) per cent of the SNR > 5 (8) peaks are matched to the same haloes. The comparison indicates minor dependence on the matching algorithm due to projection effects but we find this does not impact our main conclusions.

[2] We find that allowing a peak to be matched to a halo in the previous or subsequent shells if the haloes are close to the edge of the shell improves a small number of matches. In these cases, on one side of the edge, there is a high-valued WL peak without a corresponding massive halo. On the other side, at the same angular position, there is a massive halo without a corresponding peak. We allow a peak to be matched to haloes that are within 1 cMpc of the shell edge, which corresponds to $R_{\rm 200c}$ of a $M_{\rm 200c} \approx 10^{14}$ $M_\odot$ halo. Doubling this value does not significantly change the results.

[3] We find the results do not change if we instead match peaks to haloes based on $\Delta\theta/L_{*,\rm 50kpc}$ or $\Delta\theta/M_{*,\rm 50kpc}$, where $L_{*,\rm 50kpc}$ and $M_{*,\rm 50kpc}$ are the stellar luminosity in the GAMA r band and stellar mass in a 50 kpc aperture, respectively.

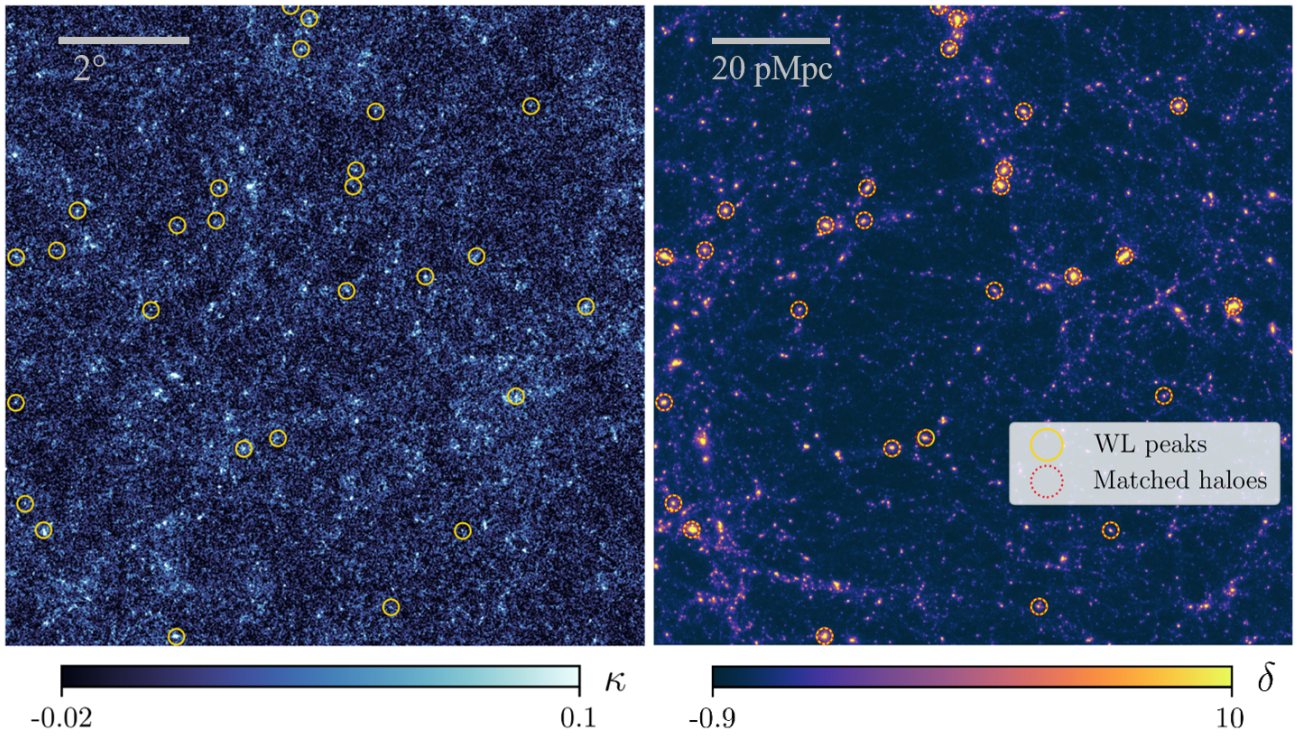


Figure 3.3: Left: $10\times$ 10 deg^2 patch of the *Euclid*-like WL convergence field of the virtual observer in L1_m9. The WL peaks are selected from this map. The SNR > 5 peaks that are assigned to the fourth shell ($z = [0.15-0.2]$) are indicated by the yellow circles. Right: projected overdensity field of the fourth shell of the same angular field, corresponding to a physical size of $\approx 110\times110$ pMpc^2. The yellow circles correspond to large overdensities. Red dotted circles show the positions of the haloes to which the peaks are matched. Most of the visibly overdense structures with no associated peak generate a peak with a lower SNR-value than the cut we apply, as illustrated in Appendix 3.C.

Fig. 3.3 illustrates the accuracy of the matching procedure. The left-hand panel shows a 10×10 square degree patch from the final WL convergence map of the virtual observer in L1_m9 from which the peaks are selected. In this panel, the SNR > 5 peaks that correspond to the fourth shell ($z = [0.15 - 0.2]$) are indicated by yellow circles. The bright peaks that do not have a circle have been assigned to other redshift shells. The right-hand panel shows the overdensity field of the fourth shell for the same angular region. The field corresponds to a physical size of $\approx 110 \times 110$ pMpc^2. Each yellow circle assigned to this shell visually corresponds to an overdense structure. The red dotted circles indicate the positions of the haloes to which the peaks have been matched. All peaks and haloes are at almost the same position, indicating the peaks are matched correctly.

Some peaks, such as the one to the left of the legend, have a small angular offset compared to the halo centre. As no other haloes are near the same position, we are still confident that the peaks have been matched to the correct haloes. Minor differences between the halo light-cone coordinates and smoothed projected WL overdensity are expected, as we apply noise and smooth the final maps and because of projection effects. Other structures seemingly equally overdense as the ones that are associated with a WL peak do not have a corresponding peak. Most of these structures cause a peak in the final WL convergence map with an SNR value below the cut-off value we apply, as illustrated in Appendix 3.C. Alternatively, a peak at that position has been assigned to a more strongly contributing structure along the line of sight. As we chose to keep the peak catalogue relatively pure, we do not aim to include these peaks in this analysis. Still, we stress that the SNR cut we applied has not been optimised and may be lowered to include more peaks, albeit at the cost of purity.

In Appendix 3.C, we split the peaks in Fig. 3.3 into the three categories considered in this work (i.e., Single shell, Noise, and Multiple shells peaks). There, we show that, generally, Noise and Multiple shells peaks are still clearly assigned to a massive halo. In that appendix, we also show the $3 <$ SNR < 5 peaks that were assigned to this shell and illustrate that most overdense structures that do not have a corresponding peak in Fig. 3.3 generate a peak with an SNR-value below the cut that we apply.

3.3.5 Observed redshifts

Until now, we have used the simulation properties to identify structures with high precision and accuracy at any point in space and time within the simulation. However, as observations typically assign a redshift based on a limited number of photometric bands, we explore how the expected associated observational uncertainties impact our results.

Euclid will rely on photometric redshifts for the bulk of the measured galaxies. The survey requirement is to measure the photometric redshifts such that

the standard deviation on the uncertainty satisfies $\sigma_z < 0.05(1+z)$ with a catastrophic failure rate of less than 10 per cent (Euclid Collaboration: Mellier et al., 2025). However, the peaks we measure correspond to the most massive clusters, as we will show in Section 3.4.1, which will likely contain bright and red objects that will be easier to identify using either the regular methodology or a red-sequence cluster identification mechanism such as redMaPPer (Rykoff et al., 2014). Therefore, assuming the general *Euclid* requirement will overestimate the redshift uncertainty on our peak sample. The assumption that the WL peaks can be associated with optical clusters using red-sequence-based cluster finders is confirmed by the recent HSC Y3 analysis of Chen et al. (2025) and Chiu et al. (2024), who find that they can match all their SNR > 5 peaks selected from aperture mass maps to an existing measurement of an optical cluster from CAMIRA (Oguri et al., 2018), redMaPPer or WHT (Wen & Han, 2015).

The *Euclid* satellite will measure photometry in 1 visual ($m_{\rm VIS}$) and 3 near-infrared (YJH) bands, but the photo-z measurements will strongly rely on ground-based photometry. The satellite's measurements will be complemented by 6 ground-based photometric measurements in the $ugrizy$ bands from a range of different facilities (Euclid Collaboration: Pocino et al., 2021; Euclid Collaboration: Mellier et al., 2025). As a more accurate estimate of the redshift uncertainty, we take as a reference the KiDS bright sample, which is a bright subsample of $\sim 10^6$ galaxies of KiDS-1000 that uses the 9-band $ugriZYJHK_s$ photometry to determine photo-zs (Kuijken et al., 2019; Bilicki et al., 2021). By comparing to the GAMA spectroscopic redshifts, they find that they assign photo-zs to the bright subsample with an accuracy of $\sigma_z = 0.018(1+z)$ up to $z = 0.5$ without catastrophic failures. Vakili et al. (2023) select a KiDS-1000 subsample using a red sequence template that includes estimating the red sequence redshifts, and by comparing to spectroscopic redshifts, they find similar accuracy and precision as the KiDS bright sample extending up to $z = 0.8$.

As the number of bands and depth used for the *Euclid* photometric redshift estimates will be similar to KiDS, we expect that the photometric redshifts for the clusters that correspond to our WL peaks will be measured with similar accuracy and precision as the KiDS bright sample. To mimic the observed redshifts, for each $z_{\rm true}$ that has been assigned to a peak, we therefore draw an observed redshift according to a normal distribution with mean μ and standard deviation σ,

$$\mathcal{N}\Big\{\mu = z_{\rm true}, \sigma_z = 0.02(1+z_{\rm true})\Big\}, \tag{3.11}$$

which is truncated at $z = 0$. We will only show the redshift distributions of all the baryonic and cosmology variations as a function of the mock photometric redshifts $z_{\rm obs}$. Still, we note that the same conclusions hold if we instead use the true redshifts or modify the redshifts with an uncertainty of $\sigma_z = 0.05(1+$

$z_{\rm true}$) and an extreme outlier fraction of 10 per cent, to the degree of the *Euclid* requirements.

3.4 Results

We now present the results of the halo peak matching by showing the WL peak distribution binned by halo mass in Section 3.4.1. We then quantify the completeness of the matched haloes in Section 3.4.2 and study the redshift distribution of the WL peaks in Section 3.4.3. We quantify the degree of numerical convergence and cosmic variance in Sections 3.4.3 and 3.4.3, respectively. Finally, we qualitatively and quantitatively compare the difference between the cosmological and baryonic impact on the peak statistics in Sections 3.4.4 and 3.4.5.

3.4.1 Peak distribution in halo mass

First, we show the results of matching the haloes to peaks by binning the peaks by halo mass. Fig. 3.4 shows the number density of WL peaks as a function of WL convergence for all redshifts combined. The top x-axis shows the SNR for the *Euclid*-like analysis. The solid black curve gives the distribution of all peaks measured in L1_m9. The SNR > 5 selection, which primarily corresponds to a single halo per peak, is binned by the halo masses of the haloes to which the peaks have been matched, as depicted using increasingly darker colours for the larger halo mass bins. The figure also shows the Noise peaks (dot-dashed) and Multiple shells peaks (dashed), as detailed in Section 3.3.2. The fraction of Noise peaks decreases with increasing WL convergence. At SNR = 5, the Noise peaks account for ~ 10 per cent of all peaks, but this quickly decreases to 0 at SNR = 10 (see also Fig. 3.2).

The bottom panel of Fig. 3.4 shows the fractional contribution of all mass bins. As expected, we find that, on average, higher-valued WL peaks correspond to more massive haloes. We find qualitatively similar results to Hamana et al. (2004); Yang et al. (2011); Liu & Haiman (2016), who found SNR > 3.5 peaks are predominantly caused by a single massive halo each. We can now study the peaks as a function of their halo mass in more detail. For $\kappa > 0.25$, we find that over 95 per cent of the peaks can be attributed to haloes with $M_{200c} > 10^{14}$ $M_\odot$. Of all SNR > 5 peaks, the majority are matched to haloes with $M_{200c} = 10^{14}$–$10^{14.5}$ $M_\odot$. In Broxterman et al. (2024), we already hypothesised that the WL peaks originate primarily from $M_{200c} > 10^{14}$ $M_\odot$ haloes based on the differences between the thermal and kinetic jet AGN variations and the comparison of the HMFs of these runs. We now confirm this hypothesis.

Liu & Haiman (2016) used observational data from the CFHTLens survey to link haloes to WL peaks. They found that each peak with SNR > 3.5 corresponds to a single halo of $M_{\rm vir} \sim 10^{15}$ $M_\odot$. Qualitatively, this agrees with our findings as

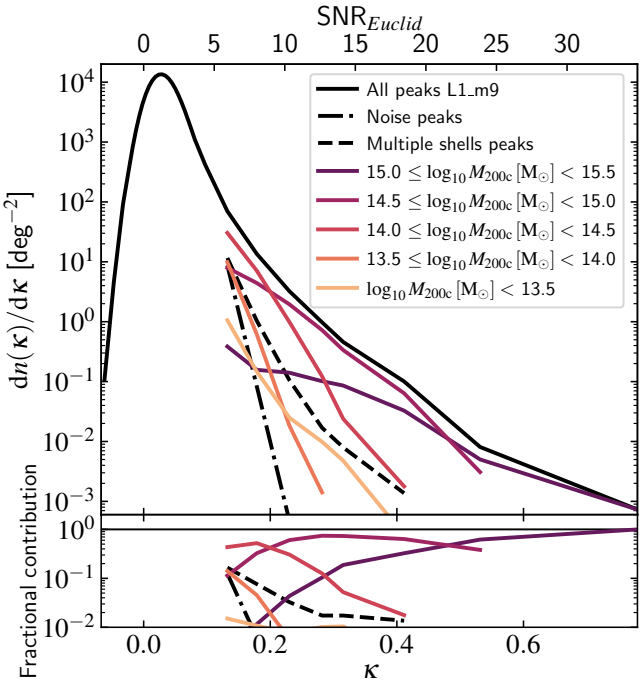


Figure 3.4: Top: number density of WL peaks as a function of WL convergence (bottom x-axis) and SNR (top x-axis) for the virtual observer in L1_m9 (solid black). Noise and Multiple shells peaks are indicated by black dashed-dotted and dashed curves, respectively. Peaks with SNR > 5 are binned by M_{200c} of the haloes to which they have been matched, with larger mass bins indicated by increasingly darker solid curves. Bottom: fractional contribution of all mass bins to the total number of haloes. The high-valued ($\kappa > 0.11$; SNR > 5) WL peaks primarily trace haloes with $M_{200c} > 10^{14}$ M$_\odot$.

their analysis probes less deep, but our highest-valued κ peaks are dominated by haloes with $M_{200c} > 10^{15}$ M$_\odot$.

3.4.2 Completeness

We define completeness as the fraction of haloes within a halo mass bin matched to at least one WL peak over the total number of haloes in that bin. Fig. 3.5 shows the completeness as a function of redshift and binned by halo mass. The solid curves correspond to haloes matched to peaks with SNR > 5, and the dashed curves include only those matched to SNR > 8 peaks. The cumulative completeness up to $z = 0.5$ and 1.0 for haloes matched to SNR > 5 peaks is given in Table 3.2. The table also includes the total number of haloes per square degree on

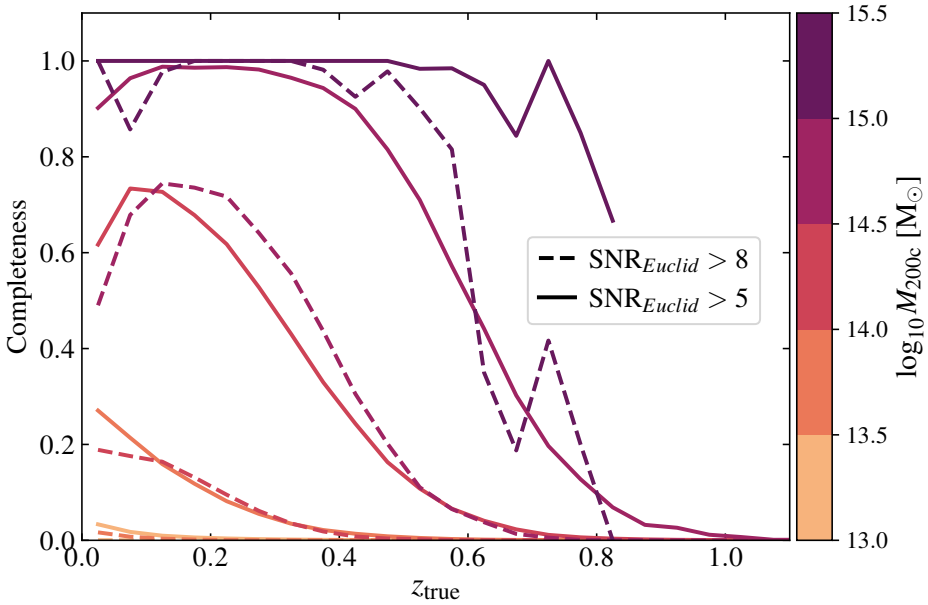


Figure 3.5: The completeness, i.e. the fraction of haloes matched to a high-valued WL convergence peak, as a function of the halo redshift. The haloes are binned by M_{200c}, and more massive halo mass bins are shown in increasingly darker colors. The solid curves correspond to the fiducial SNR > 5 sample, while the dashed curve only selects haloes matched to peaks with SNR > 8. The completeness drops for lower masses and higher redshifts. The impact of the smoothing scale on the completeness is studied in more detail in Appendix 3.A.

the halo light-cone until $z = 3$. We find excellent completeness for the most massive haloes. Of all 929 $M_{200c} > 10^{15.0}$ $M_\odot$ haloes, 98 per cent has been matched to an SNR > 5 peak. Up to $z = 0.5$, this fraction is 100 per cent. Of all 5.0×10^4 $M_{200c} = 10^{14.5} - 10^{15.0}$ $M_\odot$ haloes, 52 per cent has a corresponding SNR > 5 peak. Until $z = 0.5$, this is 93 per cent. At higher redshifts, as well as for lower masses, the completeness drops. This is expected, as lower mass objects, on average, generate weaker WL signals, higher-redshift objects have fewer source galaxies to lens, and the contribution of the WL kernel decreases at larger redshifts.

For the $10^{14} < M_{200c}[M_\odot] < 10^{15}$ bins, we predict decreasing completeness at low redshift ($z < 0.2$), which can be attributed to the fixed smoothing scale that is applied. For some of these objects, the smoothing scale is too small to achieve the optimal SNR, as they subtend a larger angular scale than is smoothed over. We illustrate this in more detail in Appendix 3.A, where we show the completeness increases for these low-redshift bins when adopting a larger smoothing scale. The most massive haloes are so massive that they are matched to a WL peak in-

Table 3.2: The completeness, i.e. the fraction of haloes matched to an SNR > 5 WL peak, assuming a smoothing scale of 1 arcmin. This peak sample has a purity of 76 per cent (see Fig. 3.2). The columns list the halo mass bin, M_{200c}; the total number of haloes per deg^2 on the entire halo light-cone until $z = 3$; completeness up to $z = 0.5$ and 1.0.

$\log_{10} M_{200c}[M_\odot]$	No. haloes [deg^{-2}]	Completeness $z < 0.5$	Completeness $z < 1.0$
13.0 – 13.5	8.9×10^2	0.002	0.001
13.5 – 14.0	1.6×10^2	0.038	0.010
14.0 – 14.5	1.9×10^1	0.378	0.136
14.5 – 15.0	1.2×10^0	0.930	0.578
15.0 – 15.5	2.3×10^{-2}	1	0.984

dependent of the smoothing scale we use. As the redshifts of the peaks are not known before detecting them, a smoothing scale has to be chosen without considering that information. The adopted smoothing scale is a trade-off between higher completeness at lower or higher redshifts for a larger or smaller smoothing scale than the fiducial value adopted in this work, respectively. We adopt a 1 arcmin scale to facilitate the comparison with Broxterman et al. (2024), but multiple scales may be combined to extract more information (Liu et al., 2015).

The dashed curves in Fig. 3.5 illustrate that the completeness drops when choosing a higher SNR threshold for the WL peaks. The increase in SNR-cut impacts the completeness of the most massive bin the least, as these haloes generally cause the highest-valued WL peaks.

Hamana et al. (2004) report their completeness in terms of the signal expected for a halo of mass similar to that in their simulation, but with an NFW profile instead of the actual profile of the halo they study. They report a completeness of 81 per cent for haloes with $\mathrm{SNR}_{\mathrm{NFW}} > 5$. In our analysis, instead, we illustrate the completeness binned by halo mass. However, at the same time, they reported an efficiency, i.e. a purity, of 82 per cent using this $\mathrm{SNR}_{\mathrm{NFW}}$ cut. We find similar, although slightly better, values based on our methodology. Combined with the purity shown in Fig. 3.2, we find that an SNR > 5 cut will result in a pure sample (76 per cent) while measuring a large number of peaks ($5.5{\times}10^4$). We repeat that extending this to lower SNR is likely possible, as we find that most Noise and Multiple shells peaks are still attributable to a massive object.

Comparison to X-ray and SZ cluster abundance inferences

Next, we compare our completeness to the state-of-the-art X-ray survey eROSITA. Ghirardini et al. (2024) use the cosmology subsample of the X-ray selected eRASS1 cluster sample in their cluster abundance inference. Their

sample, consisting of 5,263 clusters on $\sim$ 13,000 deg^2, has an estimated purity of 94 per cent (Bulbul et al., 2024). In our analysis, we find that an SNR > 8 WL peak sample, corresponding to $\sim$ 6,800 WL peaks for *Euclid* DR3, has a purity of 96 per cent, illustrating that WL peaks are competitive with state-of-the-art X-ray cluster cosmology inferences. Lowering the SNR cut to 5 results in an even greater difference in the number of probed clusters, as $N_{\rm peaks}(\mathrm{SNR} > 5) = 5.5 \times 10^4$, but this is at the cost of the purity which decreases to 76 per cent. Comparing our completeness to that of the eROSITA sample reported in Appendix D of Ghirardini et al. (2024), we see that at $z = 0.3$, their sample is $\approx$ 100 per cent complete for $M_{500c} > 10^{15.2}$ M$_\odot$. At the same redshift, as illustrated in Fig. 3.5, we find our sample (with an SNR > 5 cut) similarly complete for $M_{200c} > 10^{14.5}$ M$_\odot$. For SNR > 8, at a similar redshift, our sample has a similar completeness level for $M_{200c} > 10^{15.0}$ M$_\odot$. For $z > 0.6$, the completeness of our WL peak sample drops steeply as a consequence of the WL kernel, even for the most massive haloes.

We also compare our results to the SZ selected samples from Hilton et al. (2021). Their sample consists of 4,195 optically confirmed clusters over $\sim$ 13,000 deg^2 extending up to a redshift of 1.9. At $z = 0.5$, they report a completeness of 90 per cent down to masses of $M_{500c} > 3.8 \times 10^{14}$ M$_\odot$, which decreases for lower and higher redshifts as well as for lower masses. They expect 34 per cent of their candidates to be false positives, corresponding to a purity of 0.66. For $z > 0.6$, the completeness of their SZ-selected sample is generally better, whereas our WL sample has better completeness for lower redshifts.

To conclude, the SNR > 5 WL peak sample that Stage IV WL surveys will measure corresponds to massive clusters and has a purity and completeness equal to or better than state-of-the-art X-ray or SZ cluster abundance inferences. This especially holds for redshifts up to $z \approx 0.7$ and may be improved further by combining smoothing scales as illustrated in Appendix 3.A. At the same time, WL peaks circumvent some of the selection effects and model assumptions, as both SZ and X-ray analyses require assumptions on the parametric form and scatter in the scaling relation relating the observable to the halo mass. Compared to an X-ray or SZ cluster analysis, the limiting factor of this analysis is that the WL kernel restricts the accessible redshift range. Therefore, the redshift range that we probe does not extend to as high a redshift as is probed by X-ray or SZ cluster analyses. In Section 3.4.1 we have shown that the SNR > 5 peaks probed by stage IV WL surveys primarily trace $M_{200c} > 10^{14}$ M$_\odot$ haloes. Physically, these correspond to massive clusters of galaxies that should be visible in the optical, SZ, and X-ray (e.g. Allen et al., 2011). The analysis of this paper could, therefore, possibly be extended by including SZ-selected clusters for higher redshifts (e.g. Bocquet et al., 2024a,b).

3.4.3 Redshift distribution of SNR > 5 WL peaks

In this section, we study the redshift distribution of the SNR > 5 WL peaks. We stress that all SNR > 5 peaks, including the Noise and Multiple shells peaks, are assigned a redshift. First, we show the number density of WL peaks binned by redshift. We then investigate the degree of numerical convergence and cosmic variance. Finally, we study the redshift distributions as a function of mock photometric redshifts to quantify the impact of baryonic physics and cosmology using the different FLAMINGO 1 cGpc variations.

Fig. 3.6 illustrates the number density of WL peaks as a function of WL convergence binned by their redshift. The right-hand axis shows the number of peaks that we estimate will be detected in a $\Delta\kappa = 0.01$ bin in the final *Euclid* data release. As explained before, the redshift distribution of high-κ WL peaks depends on the evolution of structure formation and on the WL kernel set by the source galaxy redshift distribution. However, it is also sensitive to cosmology due to the comoving volume-redshift relation. Therefore, the WL peak distributions contain both geometrical and perturbative information.

Fig. 3.6 illustrates that, generally, there are no peculiar features in the κ distribution of the peaks with different redshifts other than overall amplitude differences. We find that peaks are more likely to be generated from lower redshift bins (particularly $z = [0 - 0.6]$), as this is where most of the structure has formed and the WL kernel peaks for the *Euclid*-like source distribution, independently of the WL convergence value. Only the lowest redshift bin, $z = [0 - 0.2]$ (dark blue), shows different behaviour, as it contains fewer $\kappa \lessapprox 0.15$ peaks than in the $z = [0.4 - 0.6]$ and $[0.6 - 0.8]$ bins, but for larger WL convergence values, it has more peaks. This characteristic originates from a subset of the high-mass haloes at low redshifts that generate multiple WL peaks. As discussed in the previous section, at low redshifts, the applied smoothing scale is smaller than the angular scale subtended by the most massive objects. These objects may therefore generate multiple peaks, which is reflected in Fig. 3.6 as the lowest redshift bin, compared to the $z = [0.4 - 0.8]$ bins, contains more high-valued WL peaks but fewer low-valued peaks. This behaviour may be avoided by requiring a minimal angular distance between different WL peaks entering the catalog, an effect not dissimilar to fiber collisions.

Numerical convergence

We now establish the degree of numerical convergence by comparing the redshift distributions as measured for the observers in the simulations with different resolutions and box sizes in the FLAMINGO suite. In Fig. 3.7, we show the signals measured for the L1_m8_DMO (red), L1_m9_DMO (green), L1_m10_DMO (yellow), and the mean of the eight L2p8_m9_DMO (blue) observers. The lower (m10) and higher (m8) resolution runs have, respectively, 8 times fewer and more

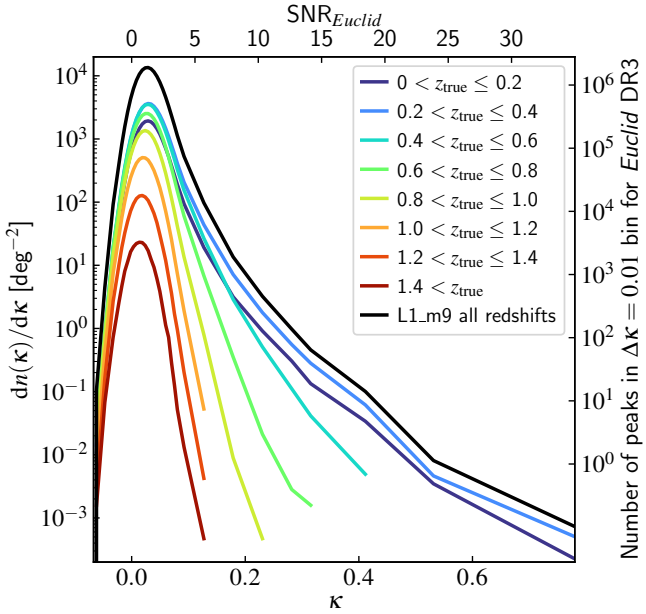


Figure 3.6: Number density of WL peaks as a function of WL convergence binned by redshift for L1_m9. The redshift is assigned by identifying the lensing plane that contributes most to the final WL peak convergence value (Equation 3.7). The redshift distribution is determined by the WL kernel, which is set by the source redshift distribution, and the evolution of structure formation. The distribution peaks at z = [0.2-0.4] as expected from the WL kernel peaking at $z \approx 0.375$. The right axis indicates the number of WL peaks in a $\Delta\kappa = 0.01$ bin for *Euclid* DR3.

particles than m9 in the same volume and with the same initial conditions. We illustrate the redshift distributions as the number density of the redshifts of WL peaks $dn(z)/dz$, i.e. the counts per redshift bin over the bin size ($\Delta z = 0.05$).

The distributions peak at $z \approx 0.3$ but are moderately flat across the $z = 0.2-0.4$ range. Whereas the WL kernel already peaks at $z \approx 0.375$, more massive structures continue to form at low z, effectively pulling the distribution to lower redshifts than if we had only considered the WL kernel. The bottom panel shows the ratio compared to L1_m9_DMO, with the gray band indicating the Poisson error. The observers in the L1 boxes are located at the same position, and these boxes have the same initial phases. Therefore, comparing these three variations directly shows the degree of numerical convergence with the particle mass and spatial resolution. We find a similar degree of convergence as the number density of WL peaks in Broxterman et al. (2024). The m10 run shows a systematic offset of a few per cent for $z = [0, 0.7]$. As the simulations are initiated with the same phases

and based on the convergence of the HMFs of these runs (see fig. 19 of Schaye et al. (2023)), we expect the same massive haloes to exist in the different variations. Because the WL peak sample primarily probes these massive halos (see Fig. 3.4), we do expect that the SNR > 5 haloes probed by L1_m9_DMO should still be well resolved in L1_m10_DMO. The small but systematic suppression of the L1_m10_DMO signal shows that the convergence is slightly worse than the convergence of the HMF, which may be a consequence of halo profiles being less resolved and the peak height in the lower resolution simulation not making the SNR > 5 cut. The convergence between L1_m9 and L1_m8 is much better. There is a 5 per cent difference for the lowest redshift bin, which decreases and remains within 1 per cent up to $z = 0.7$.

The L2p8_m9 and L1_m9 boxes differ in the realization of the initial conditions. Therefore, they are also impacted by cosmic variance and illustrate the convergence with the volume probed. The convergence with box size is slightly worse than with resolution but still within 3 per cent for $z = 0.1 - 0.8$ and within 1 per cent until for $z = 0.25 - 0.50$, where the distribution peaks. In general, the difference is similar to the Poisson error for L1_m9.

Whereas Fig. 3.7 shows the numerical convergence for the signal measured in the DMO boxes, we note that the same conclusions apply to the signals measured in the full hydrodynamical simulations.

Cosmic variance

Using the eight different observers in the L2p8_m9_DMO box, we test the impact of cosmic variance on our statistics. Again, we show the results for the observers in the DMO run, but we find that the level of difference is the same for the observers in the hydrodynamical simulation. Fig. 3.8 shows the redshift distribution of SNR > 5 WL peaks for the eight virtual observers. The distributions depend on the specific orientation of the individual observer with respect to the most massive haloes in the simulation. The bottom panel shows the ratio of the distribution for each observer to the mean of the eight observers. The degree of cosmic variance is of the same level as the Poisson noise, as shown in the bottom panel of Fig. 3.7. The effect is less than 1 per cent for $z = 0.1 - 0.6$.

3.4.4 Baryonic and cosmology impact

Next, we compare the distributions of the virtual observers in the 1 cGpc simulations with different levels of baryonic feedback or cosmologies. We emphasize that all observers in the L1 boxes are located at the same position and that the simulations were initiated with the same initial phases. All SNR > 5 peaks in the final WL convergence maps are assigned the redshift of the lens plane that contributes most to the WL peak convergence value. At each lens plane, the WL peaks are matched to individual haloes and we assign the redshift of the haloes

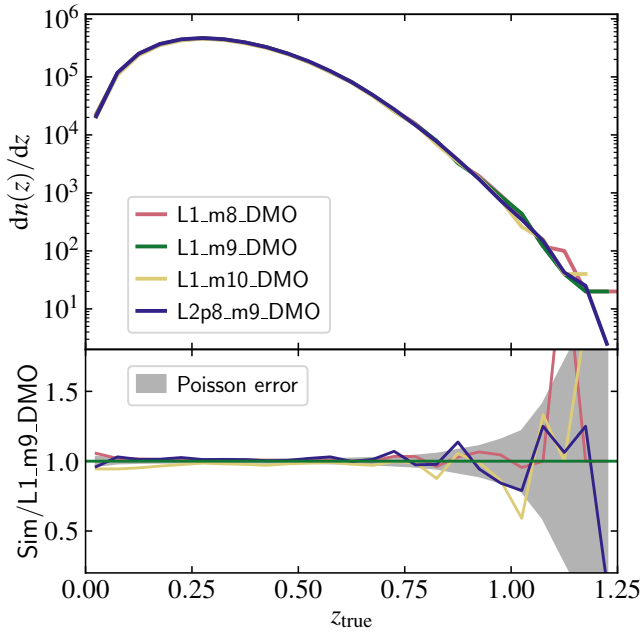


Figure 3.7: Top: number density of the redshifts of SNR > 5 WL peaks for L1_m8_DMO (red), L1_m9_DMO (green), L1_m10_DMO (yellow), and the mean of the eight observers in L2p8_m9_DMO (blue). Bottom: Ratio to L1_m9_DMO. The gray band indicates the Poisson error. The agreement between the variations shows good convergence of the L1_m9 simulations with box size and resolution.

to the peaks.[4] We then adjust the redshifts to account for the expected uncertainty for the massive objects we probe (equation 3.11). In Fig. 3.9, we show the ratio of the redshift distributions ($\mathrm{d}n(z)/\mathrm{d}z$) of the WL peaks of all L1 variations (top panels) and their ratio with respect to L1_m9 (bottom panels) as a function of the perturbed observed redshifts z_{obs}. However, we have found that the same conclusions hold when not including any observational redshift uncertainty or catastrophic outliers to the degree of the *Euclid* requirements. We bin the distributions using $\Delta z = 0.1$.

[4]The FLAMINGO halo light-cones for L1_m9_DMO and the decaying dark matter variations are currently unavailable. We, therefore, show the results of these variations using the central redshifts of the shells to which the peaks have been assigned. As we only show relative differences to L1_m9, we also restrict the L1_m9 shell redshifts when comparing to L1_m9_DMO, PlanckDCDM12 and PlanckDCDM24. We confirmed that the ratio to L1_m9 is the same for the simulation variations that do have halo light-cones.

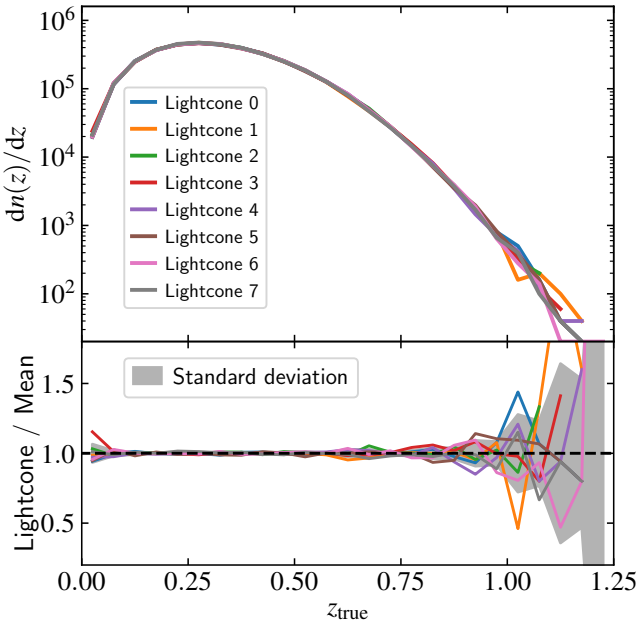


Figure 3.8: Top: number density of the redshifts of SNR > 5 WL peaks for the eight virtual observers in L2p8_m9_DMO. Bottom: ratio to the mean redshift distribution viewed by the eight observers. The standard deviation in the bottom panel, shown by the gray-shaded region, estimates the effect of cosmic variance.

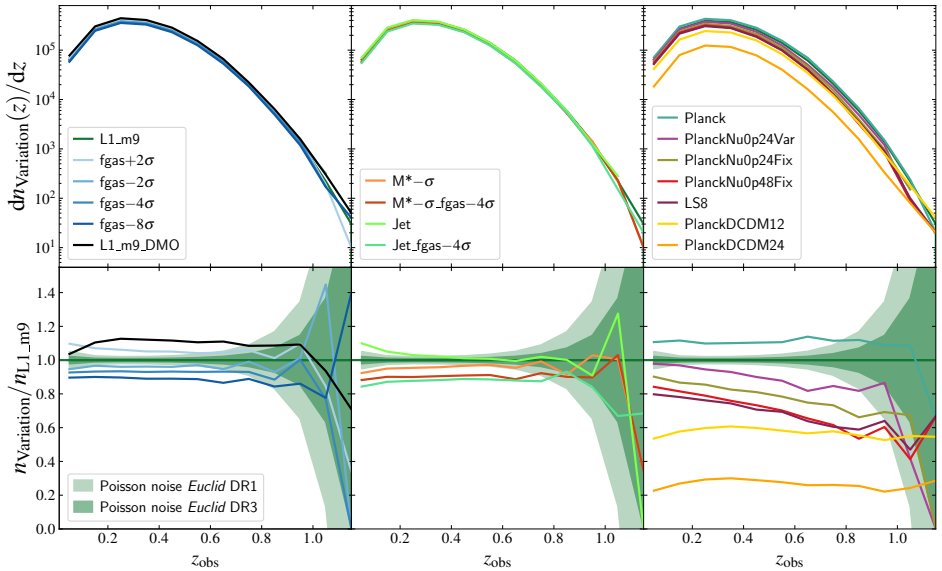


Figure 3.9: Top: number density of the redshifts of SNR > 5 WL peaks for all L1_m9 FLAMINGO variations. Bottom: Ratio of the redshift distribution of the WL peak number density relative to the prediction for L1_m9 (dark green curve). The left panels includes the variations varying the gas fractions in clusters, with more suppressed gas fractions indicated by an increasingly darker blue colour and the DMO signal in black. The middle panels shows the remaining baryonic variations with a different AGN model ('Jet') or a suppressed galaxy stellar mass function ('M*−σ'). The right panels shows the cosmology variations, including multiple variations using more massive neutrinos or decaying dark matter ('DCDM'). The shaded green regions in the bottom panels indicate the expected Poisson uncertainty on the *Euclid* DR1 and DR3 measurements. The baryonic variations (left and central panels) show little change in the shape of the redshift distribution compared to the fiducial model, while some of the cosmology variations, particularly the more massive neutrino variations and LS8, show a different shape redshift evolution. The cosmology variations also show a greater overall offset than the baryonic variations. The shape of the redshift distribution of high-SNR WL peaks is insensitive to the intensity of baryonic feedback, but is sensitive to cosmological parameters, making it potentially valuable for cosmological inference.

The green shaded region in the bottom panels indicates the Poisson error for the expected number of peaks in *Euclid* data release 1 (DR1, $\sim 2,500$ deg^2) and for the final data release (DR3, $\sim 14,000$ deg^2). The Poisson errors overestimate the differences between the simulation variations as all L1 variations were initiated with the same initial phases, and we thus expect the simulations to be correlated.

The left and central panels show the predictions for L1_m9_DMO and the baryonic feedback variations. These variations do not show a different redshift evolution from the fiducial L1_m9 model as all curves in the bottom panels are close to horizontal, which even holds for the strongest feedback model, fgas−8σ. The systematic offsets in amplitude between the curves stem from the fact that the total number of peaks above the SNR cut differs between the simulations. In simulations with stronger feedback, an individual halo that acts as a lens has less mass (e.g. Velliscig et al., 2014). Therefore, the WL convergence value of the peak corresponding to that halo will be lower. As we apply a fixed SNR cut, we expect the total number of peaks to be smaller in simulation with stronger feedback, as was shown in Broxterman et al. (2024) and is reflected in the vertical offsets in Fig. 3.9.

After correcting for the amplitude difference, all deviations from the shape of the fiducial redshift evolution are within a couple of per cent, though the amplitude of the distributions varies by ~ 10 per cent. The most distinct baryonic variations are L1_m9_DMO and Jet. The DMO prediction shows a minor decrease for $z < 0.2$. However, when quantifying the difference between the distributions in the next Section, we still find the two distributions are statistically similar. The Jet model was calibrated to the same observables as L1_m9, and the only difference is the subgrid implementation of AGN feedback. At $z = 0$, Jet shows some enhancement compared to L1_m9, which decreases for higher z, and at $z > 0.3$, the kinetic and thermal AGN feedback variations give the same result. The comparison indicates some dependence on the simulation's AGN subgrid physics model. Comparing the $z = 0$ HMFs of L1_m9 and Jet as shown in Fig. 20 of Schaye et al. (2023), we observe that the Jet model, compared to L1_m9, has a few per cent more haloes of mass $M_{200c} = 10^{13} - 10^{14}$ M$_\odot$. This excess of haloes could explain the overabundance of WL peaks we observe at low z.

The lack of differences in the redshift evolution between the baryonic feedback variations is consistent with the interpretation that the redshift distribution of the high-SNR WL peaks is set by the formation time of massive haloes and is insensitive to the strength of the baryonic feedback that manifests itself primarily only once a halo has formed and has accreted sufficient mass. The strength of the feedback does not impact the time at which a halo has formed but only whether it creates a WL signal significant enough to be included in the SNR > 5 sample, as reflected in the amplitude differences as shown in Figure 3.9. This reasoning still holds if baryonic feedback is redshift dependent or if haloes of the same mass experience different baryonic feedback at different redshifts, provided

the WL peaks correspond to single massive objects, as we have shown is the case, and the baryonic evolution is independent of cosmology, which is to first order the case for WL peaks measured in the limited number of FLAMINGO variations at different cosmologies. As the shape of the L1_m9_DMO redshift evolution is consistent with that for the hydrodynamical simulations assuming the same cosmology, the redshift evolution of the WL peak sample studied in this work may be modeled with N-body simulations only.

Instead, if we focus on the cosmology variations in the right panels, we see greater differences in amplitude and evolution. We distinguish two subgroups in these panels. The first group, consisting of the neutrino mass variations and LS8, shows a different redshift evolution compared to the runs using the fiducial cosmology. Neutrinos act as a form of hot DM, thereby suppressing structure formation (e.g. Lesgourgues & Pastor, 2006). This is reflected in the lower value of σ_8, as shown in the next to last column in Table 3.1. LS8, which has a lower value of σ_8 by design, and the higher neutrino mass variations have a lower value of σ_8 than the fiducial cosmology. Consequently, there are fewer massive haloes in these simulations. This difference in the number of massive clusters in these simulations, compared to L1_m9, increases with redshift. This is also reflected in the redshift distributions, as those belonging to the more massive neutrino and LS8 variations are increasingly more suppressed at higher z.

The second group, consisting of the Planck and decaying dark matter models, shows no strong difference in the redshift evolution compared to L1_m9. The *Planck* cosmology has a 3.3 per cent higher value of Ω_m and only a 0.6 per cent lower value of σ_8. Considering the HMF, larger values of Ω_m increase the overall amplitude of the HMF, whereas a lower σ_8 moves the exponential drop to lower halo masses (e.g. Xhakaj et al., 2024). Assuming all peaks emanate from a single halo, we expect the Planck and L1_m9 (i.e. DES Y3 3 × 2pt + All Ext.) to differ primarily in the total number of haloes as the difference in Ω_m is a factor 5 greater than the difference in σ_8. This is reflected in the redshift distributions, as the Planck model has more peaks overall, but there is no apparent difference in the redshift evolution. However, more variations should be compared to investigate the impact of different cosmological parameters on the WL peak redshift evolution.

In the DCDM variations, each dark matter particle, independent of its velocity or local density, loses part of its mass as set by the decay rate indicated in the final column of Table 3.1. The half-life of these models is relatively short compared to values that have already been ruled out by geometrical constraints from the BAO (e.g. Aubourg et al., 2015), but the variations can still be used to understand the impact of decaying dark matter qualitatively. As the DM decays, the total matter density decreases, as indicated in the 9th column of Table 3.1, and the growth of structure formation slows down (e.g. Aoyama et al., 2014; Enqvist et al., 2015). This is reflected in Fig. 3.9 as the suppression of the PlanckDCDM12 and 24

models is, respectively, a factor 2 and 3 greater than for the other variations. The decaying dark matter models are variations on the Planck model, so we expect them to converge back to Planck at high redshift. Whereas visually the DCDM models do not show large differences from L1_m9, in the next Section, we will quantify the difference in redshift evolution; we find that the DCDM variations are still discrepant to the degree that we expect *Euclid* to be able to discriminate between them based on the shape of the redshift distribution of SNR > 5 peaks. The mass loss in the DCDM becomes increasingly more apparent at low redshift. This is reflected in the DCDM signals as the suppression increases for $z < 0.4$, which can be interpreted as a larger fraction of low−z lenses that lose enough mass to drop below the SNR threshold.

The comparison illustrates that the redshift evolution of WL peaks can constrain parts of the cosmological parameter space; as we have illustrated, it is sensitive to the neutrino mass. However, it also shows that changes in cosmology may result in degenerate signatures when using only WL peaks. For example, the most massive neutrinos run and LS8 are practically indistinguishable.

In Broxterman et al. (2024), we showed that even though the impact of cosmology on the number density of the WL peaks is greater than that of realistic baryonic feedback, the two signatures are degenerate. In this work, we have shown that changes in cosmology impact the redshift evolution of the number density of high-SNR WL peaks more than baryonic feedback is expected to do. Combining the redshift and WL convergence distributions can, therefore, not only help constrain cosmology, but also simultaneously calibrate baryonic feedback. The evolution of the redshift distribution of the WL peaks will yield the cosmology, while the number density of WL convergence peaks is sensitive to the baryonic feedback strength. However, as indicated by the Planck variation, not all cosmologies change the redshift distribution of the peaks. Baryonic feedback and changes in cosmology are thus still degenerate in certain parts of the parameter space. From the variations we explored, this seems to be the case for $\Omega_{\rm m}$ and baryonic feedback strength leading to lower gas fractions in low-z clusters. However, more variations that jointly vary baryonic physics and cosmology should be compared in order to study this in more detail.

Ultimately, baryonic physics could be calibrated jointly while constraining cosmology by combining different WL statistics (see e.g. Semboloni et al., 2013), that are sensitive in different ways to cosmology and baryonic physics. This may be extended by including (the cross-correlation with) additional probes to inform on the strength of baryonic, such as SZ or X-ray (see e.g. McCarthy et al., 2023, 2025; Bigwood et al., 2024). In this section, we have argued that WL peaks will be useful in furthering this goal.

3.4.5 Cosmology with the redshift of WL peaks

Next, we quantify the differences between the simulation variations to L1_m9 to determine the feasibility of using WL peaks to constrain cosmology with Stage IV WL surveys. As we are interested in the shape of the distribution, as opposed to its amplitude, we use the KolmogorovSmirnov test. The non-parametric test evaluates the null hypothesis that two distributions are equal by comparing their cumulative density functions. For each of the L1 variations, we randomly draw the number of peaks corresponding to our *Euclid* DR3 estimate (i.e. corresponding to 14,000 deg^2 instead of the full light-cone) and compare this to the L1_m9 distribution, which yields a p-value describing the hypothesis of the distributions being identical. We randomly draw a WL peak sample 1000 times and report the mean of the p-values for the WL convergence and redshift number densities in the second and third columns of Table 3.3, respectively. Entries with a greater value are more consistent with the fiducial simulation. In bold, we highlight all entries with a p-value lower than 0.1 to illustrate variations whose distribution differs most from L1_m9. The differences in the κ distributions are generally greater than those in the redshift distributions. Similarly, the differences induced by the changes in cosmology are greater than those between the baryonic feedback variations.

Focusing first on the difference in the κ distributions, we expect all cosmology variations to result in number density distributions that are discrepant with L1_m9 to the degree that should be measurable with the *Euclid* DR3 sample. At the same time, we see that some of the baryonic variations, specifically the simulations calibrated to $f_{\rm gas} - 8\sigma$, M*–σ_fgas–4σ and Jet_fgas–4σ, also result in similar differences. This means the baryonic feedback should be understood and modeled accurately before the κ number density of WL peaks can be exploited to constrain cosmology.

In contrast, based on the bold-valued entries for the p_z-values, we generally find that the baryonic feedback variations are consistent with L1_m9, whereas the cosmology variations show great differences as all but Planck and the DCDM variations have p-values lower than 10^{-5}. Only one baryonic feedback model, the Jet model, is found to have a redshift distribution with a shape statistically different from L1_m9. This is not the case for the Jet model with stronger feedback, Jet_fgas–4σ. The origin of this difference is unclear, but may indicate some dependence on the simulation's AGN subgrid model.

If we carry out a similar test to compare fgas–8σ to PlanckNu0p24Var we find that the corresponding p-values are $p_\kappa = 0.48$ and $p_z = 6.5 \times 10^{-5}$, i.e. the shape of their distributions in WL convergence is indistinguishable while their redshift distributions are different. As these variations differ in terms of both cosmology and baryonic feedback, the comparison highlights the degeneracy between cosmology and baryonic feedback in the number density in WL convergence, which

Table 3.3: p-value of the KS test of the null hypothesis that the shape of the number density as a function of WL convergence (p_κ) or redshift (p_z) of the SNR > 5 WL peak distribution measured in a simulation variation is identical to L1_m9. A high p-value indicates that the two distributions are statistically indistinguishable. Values with a p-value lower than 0.1 are indicated in bold. The shape of the WL convergence number densities differs between cosmology and baryonic variations. The shape of the redshift distribution depends primarily on cosmology and, to a much smaller degree, on baryonic feedback.

Identifier	p_κ	p_z
L1_m9_DMO	0.43	0.18
fgas+2σ	0.38	0.16
fgas−2σ	0.47	0.64
fgas−4σ	0.16	0.67
fgas−8σ	$\mathbf{7.9 \times 10^{-3}}$	0.18
M*−σ	0.38	0.33
M*−σ_fgas−4σ	$\mathbf{1.8 \times 10^{-2}}$	0.62
Jet	0.26	$\mathbf{6.6 \times 10^{-3}}$
Jet_fgas−4σ	$\mathbf{5.4 \times 10^{-3}}$	0.45
Planck	$\mathbf{3.4 \times 10^{-2}}$	0.56
PlanckNu0p24Var	$\mathbf{7.2 \times 10^{-3}}$	$\mathbf{< 10^{-5}}$
PlanckNu0p24Fix	$\mathbf{< 10^{-5}}$	$\mathbf{< 10^{-5}}$
PlanckNu0p48Fix	$\mathbf{< 10^{-5}}$	$\mathbf{< 10^{-5}}$
PlanckDCDM12	$\mathbf{< 10^{-5}}$	$\mathbf{3.7 \times 10^{-2}}$
PlanckDCDM24	$\mathbf{< 10^{-5}}$	$\mathbf{1.2 \times 10^{-3}}$
LS8	$\mathbf{< 10^{-5}}$	$\mathbf{< 10^{-5}}$

may be broken by including the redshift information.

To conclude, the p_z values confirm the result that the redshift distribution of WL peaks depends primarily on cosmology and, to a much smaller degree, on baryonic feedback strength. Therefore, the redshift distribution of high-valued SNR peaks provides an excellent probe to help constrain cosmology complementary to the commonly used peak height distribution.

3.5 Summary and Conclusions

In this paper, we have studied the cosmological potential of the redshift distribution of the expected measurement of high-κ (SNR > 5) WL peaks for a Stage IV WL survey using the FLAMINGO cosmological hydrodynamical simulation suite. We measured the contribution to the WL convergence value along the line of sight for WL peaks that were selected from integrated maps, and we assigned

the redshift of the most contributing overdensity to each peak. At the peak redshift, we match peaks to haloes to determine the origin of the WL peaks. We find that high-SNR peaks (SNR > 5) originate primarily from single massive haloes with $M_{200c} > 10^{14}$ $M_\odot$ (Fig. 3.2) and that more massive haloes preferentially cause higher−κ peaks (Fig. 3.4). We estimate Stage IV WL surveys will detect 5.5×10^4 (6.8×10^3) SNR > 5 (8) WL peaks with a purity, i.e. fraction belonging to a single massive halo along the line of sight, of 76 (96) per cent. For haloes with $M_{200c} > 10^{14.5}$ $M_\odot$, we find a completeness, i.e. the percentage of haloes that is matched to a peak in our SNR > 5 peak sample, of 93 (59) per cent up to $z = 0.5$ (1) (Fig. 3.5). The purity and completeness are competitive with state-of-the-art X-ray and SZ cluster abundance inferences. The paper has highlighted the possibility of using the WL peaks measured by upcoming Stage IV WL surveys. In the future, the distribution of the number density of WL peaks in WL convergence and redshift may be studied as a function of cosmology, as is done commonly for the HMF.

By matching each WL peak to individual haloes and by assigning a redshift, we have studied the impact of baryonic physics and cosmology on the redshift evolution of the number density of WL peaks. We have shown that, within the FLAMINGO simulation suite, the changes in baryonic feedback strength, as parameterized by the gas fraction in clusters and/or the $z = 0$ galaxy stellar mass function, do not have a substantial impact on the redshift evolution of the sample of high-SNR WL peaks that will be measured by Stage IV WL surveys (Fig. 3.9). This includes models where the gas fractions in low−z clustered are lowered by 8σ compared to the fiducial model and a DMO+ν simulation without baryons.

In contrast, changes in cosmology can have a strong impact on the redshift distribution of high-SNR WL peaks. From the variations we have explored, we find the redshift distribution is particularly sensitive to the neutrino mass, which impacts structure formation and the value of σ_8. However, more variations that jointly vary cosmology and baryonic physics should be carried out to further quantify which regions of cosmology and baryonic physics can be effectively constrained using the redshift distribution of WL peaks. Combining the redshift distribution of WL peaks with their number density as a function of the WL convergence can help simultaneously calibrate baryonic feedback and constrain cosmology, as the two statistics show different dependencies on baryonic physics. The constraining power can be increased by including other higher-order WL statistics or by including (cross-correlations with) other observables, such as SZ or X-ray.

Acknowledgements

RK acknowledges support by research programme Athena 184.034.002 from the Dutch Research Council (NWO). VJFM acknowledges support by NWO through

the Dark Universe Science Collaboration (OCENW.XL21.XL21.025). 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. We have relied on the numerical packages MATPLOTLIB (Hunter, 2007), HEALPY (Zonca et al., 2019), NUMPY (Harris et al., 2020), SWIFTSIMIO (Borrow & Borrisov, 2020), and ASTROPY (Astropy Collaboration et al., 2022) for the analysis and plotting.

3

Data Availability

The data supporting the figures in this article are available upon reasonable request to the corresponding author.

References

Abbott T. M. C., et al., 2022, Phys. Rev. D, 105, 023520
Aihara H., et al., 2018, PASJ, 70, S8
Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Amon A., et al., 2023, MNRAS, 518, 477
Aoyama S., Sekiguchi T., Ichiki K., Sugiyama N., 2014, J. Cosmology Astropart. Phys., 2014, 021
Applegate D. E., et al., 2014, MNRAS, 439, 48
Asgari M., et al., 2021, A&A, 645, A104
Astropy Collaboration et al., 2022, ApJ, 935, 167
Aubourg É., et al., 2015, Phys. Rev. D, 92, 123516
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Benisty D., 2021, Physics of the Dark Universe, 31, 100766
Bernardeau F., van Waerbeke L., Mellier Y., 1997, A&A, 322, 1
Bigwood L., et al., 2024, MNRAS, 534, 655
Bilicki M., et al., 2021, A&A, 653, A82
Blas D., Lesgourgues J., Tram T., 2011, J. Cosmology Astropart. Phys., 2011, 034
Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, MNRAS, 456, 2361
Bocquet S., et al., 2024a, Phys. Rev. D, 110, 083509
Bocquet S., et al., 2024b, Phys. Rev. D, 110, 083510
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Borrow J., Borrisov A., 2020, Journal of Open Source Software, 5, 2430
Borrow J., Schaller M., Bower R. G., Schaye J., 2022, MNRAS, 511, 2367
Broxterman J. C., et al., 2024, MNRAS, 529, 2309
Bulbul E., et al., 2024, A&A, 685, A106
Cacciato M., van den Bosch F. C., More S., Li R., Mo H. J., Yang X., 2009, MNRAS, 394, 929
Chaikin E., Schaye J., Schaller M., Benítez-Llambay A., Nobels F. S. J., Ploeckinger S., 2023, MNRAS, 523, 3709
Chen K.-F., et al., 2025, The Open Journal of Astrophysics, 8, 2
Chiu I. N., et al., 2024, The Open Journal of Astrophysics, 7, 90
Chua K. T. E., Pillepich A., Vogelsberger M., Hernquist L., 2019, MNRAS, 484, 476
Coulton W. R., Liu J., McCarthy I. G., Osato K., 2020, MNRAS, 495, 2531
Courtin J., Rasera Y., Alimi J. M., Corasaniti P. S., Boucher V., Füzfa A., 2011, MNRAS, 410, 1911
DESI Collaboration et al., 2025, J. Cosmology Astropart. Phys., 2025, 021
Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431
Dark Energy Survey and Kilo-Degree Survey Collaboration et al., 2023, The Open Journal of Astrophysics, 6, 36
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
Davies C. T., Harnois-Déraps J., Li B., Giblin B., Hernández-Aguayo C., Paillas E., 2024, MNRAS, 533, 3546
Debackere S. N. B., Hoekstra H., Schaye J., 2022, MNRAS, 515, 6023
Dietrich J. P., Hartlap J., 2010, MNRAS, 402, 1049
Elbers W., Frenk C. S., Jenkins A., Li B., Pascoli S., 2021, MNRAS, 507, 2614

Elbers W., et al., 2025, MNRAS,
Enqvist K., Nadathur S., Sekiguchi T., Takahashi T., 2015, J. Cosmology Astropart. Phys., 2015, 067
Euclid Collaboration: Ajani V., Baldi M., Barthelemy A., et al., 2023, A&A, 675, A120
Euclid Collaboration: Blanchard A., Camera S., Carbone C., et al., 2020, A&A, 642, A191
Euclid Collaboration: Martinet N., Schrabback T., Hoekstra H., et al., 2019, A&A, 627, A59
Euclid Collaboration: Mellier Y., Abdurro'uf Acevedo Barroso J., et al., 2025, A&A, 697, A1
Euclid Collaboration: Pocino A., Tutusaus I., Castander F. J., et al., 2021, A&A, 655, A44
Forouhar Moreno V. J., Helly J., McGibbon R., Schaye J., Schaller M., Han J., Kugel R., 2025, arXiv e-prints, p. arXiv:2502.06932
Fumagalli A., Costanzi M., Saro A., Castro T., Borgani S., 2024, A&A, 682, A148
Gatti M., et al., 2023, Monthly Notices of the Royal Astronomical Society: Letters, 527, L115
Ghirardini V., et al., 2024, A&A, 689, A298
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Grandón D., Marques G. A., Thiele L., Cheng S., Shirasaki M., Liu J., 2024, Phys. Rev. D, 110, 103539
Hamana T., Takada M., Yoshida N., 2004, MNRAS, 350, 893
Hamana T., Sakurai J., Koike M., Miller L., 2015, PASJ, 67, 34
Han J., Cole S., Frenk C. S., Benitez-Llambay A., Helly J., 2018, MNRAS, 474, 604
Harnois-Déraps J., et al., 2024, MNRAS, 534, 3305
Harris C. R., et al., 2020, Nature, 585, 357
Hilbert S., et al., 2020, MNRAS, 493, 305
Hildebrandt H., et al., 2017, MNRAS, 465, 1454
Hilton M., et al., 2021, ApJS, 253, 3
Hoekstra H., Jain B., 2008, Annual Review of Nuclear and Particle Science, 58, 99
Hunter J. D., 2007, Computing in Science and Engineering, 9, 90
Huško F., Lacey C. G., Schaye J., Schaller M., Nobels F. S. J., 2022, MNRAS, 516, 3750
Kacprzak T., et al., 2016, MNRAS, 463, 3653
Kaiser N., 1992, ApJ, 388, 272
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Kratochvil J. M., Haiman Z., May M., 2010, Phys. Rev. D, 81, 043519
Kugel R., et al., 2023, MNRAS, 526, 6103
Kugel R., Schaye J., Schaller M., McCarthy I. G., Braspenning J., Helly J. C., Forouhar Moreno V. J., McGibbon R. J., 2024, MNRAS, 534, 2378
Kugel R., Schaye J., Schaller M., Moreno V. J. F., McGibbon R. J., 2025, MNRAS,
Kuijken K., et al., 2019, A&A, 625, A2
LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201
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
Li Z., Liu X., Fan Z., 2023, MNRAS, 520, 6382
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
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Lu T., Haiman Z., 2021, MNRAS, 506, 3406
Mana A., Giannantonio T., Weller J., Hoyle B., Hütsi G., Sartoris B., 2013, MNRAS, 434, 684
Martinet N., et al., 2018, MNRAS, 474, 712
Martizzi D., Mohammed I., Teyssier R., Moore B., 2014, MNRAS, 440, 2290
McCarthy I. G., Le Brun A. M. C., Schaye J., Holder G. P., 2014, MNRAS, 440, 3645
McCarthy I. G., Bird S., Schaye J., Harnois-Deraps J., Font A. S., van Waerbeke L., 2018, MNRAS, 476, 2999
McCarthy I. G., et al., 2023, MNRAS, 526, 5494
McCarthy I. G., et al., 2025, MNRAS, 540, 143

Medezinski E., et al., 2018, PASJ, 70, S28
Nagai D., Vikhlinin A., Kravtsov A. V., 2007, ApJ, 655, 98
Oguri M., Miyazaki S., 2025, Proceedings of the Japan Academy, Series B, 101, 129
Oguri M., et al., 2018, PASJ, 70, S20
Philcox O. H. E., Ivanov M. M., 2022, Phys. Rev. D, 105, 043517
Planck Collaboration et al., 2020a, A&A, 641, A6
Planck Collaboration et al., 2020b, A&A, 641, A8
Ploeckinger S., Schaye J., 2020, MNRAS, 497, 4857
Reid B., et al., 2016, MNRAS, 455, 1553
Reiprich T. H., Böhringer H., 2002, ApJ, 567, 716
Riess A. G., Casertano S., Yuan W., Bowers J. B., Macri L., Zinn J. C., Scolnic D., 2021, ApJ, 908, L6
Rykoff E. S., et al., 2014, ApJ, 785, 104
Schaller M., et al., 2024, MNRAS, 530, 2378
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Schaye J., et al., 2023, MNRAS, 526, 4978
Scoccimarro R., 2004, Phys. Rev. D, 70, 083007
Secco L. F., et al., 2022, Phys. Rev. D, 105, 023515
Semboloni E., Hoekstra H., Schaye J., 2013, MNRAS, 434, 148
Shan H., et al., 2018, MNRAS, 474, 1116
Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
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
Tröster T., et al., 2020, A&A, 633, L10
Upadhye A., et al., 2024, MNRAS, 529, 1862
Vakili M., et al., 2023, A&A, 675, A202
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
Velliscig M., et al., 2015, MNRAS, 453, 721
Wang L., Steinhardt P. J., 1998, ApJ, 508, 483
Weiss A. J., Schneider A., Sgier R., Kacprzak T., Amara A., Refregier A., 2019, J. Cosmology Astropart. Phys., 2019, 011
Wen Z. L., Han J. L., 2015, ApJ, 807, 178
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., 2024, MNRAS, 530, 4203
Yang X., Kratochvil J. M., Wang S., Lim E. A., Haiman Z., May M., 2011, Phys. Rev. D, 84, 043529
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, Journal of Open Source Software, 4, 1298
van Waerbeke L., 2000, MNRAS, 313, 524

Appendix

3.A Smoothing

In this Appendix, we study the impact of the adopted smoothing scale on the halo peak matching. The fiducial smoothing scale adopted in this work is 1 arcmin. In Fig. 3.10, we show the results of the completeness of the halo peak matching for three additional smoothing scales: 0.5 ′ (upper left), 1 ′ (upper right), 3 ′ (lower left), and 10 ′ (lower right). The curves corresponding to 1 ′ smoothing are repeated as thin lines in each panel.

For each smoothing scale, the figure shows the completeness, i.e. the fraction of haloes matched to a peak (with SNR > 5) within a halo mass bin. As already mentioned in Section 3.4.2, the small drop in completeness at low redshift for the $M_{200c} = 10^{14} - 10^{15}$ $M_\odot$ mass bins is the result of the smoothing scale being too small to achieve the optimal SNR for these objects. At $z = 0.1$, the smoothing scale of 1 ′ corresponds to a physical scale of ≈ 110 kpc, which is an order of magnitude smaller than the viral radius of $M_{200c} \geq 10^{14}$ $M_\odot$ haloes.

This smoothing scale is, therefore, too small to achieve the optimal SNR for these haloes. This is illustrated by the fact that the completeness of the same halo mass bins is larger when 3 ′ smoothing has been applied. The most massive haloes ($M_{200c} \geq 10^{15}$ $M_\odot$) do not suffer from this problem, as they are so massive that independent of the smoothing scale, they will cause a large enough WL peak to be detected. A larger smoothing scale than the fiducial 1 ′ will allow the detection of more low-z objects as this scale corresponds better to the angular size of the objects we are probing. However, simultaneously, the completeness at higher z drops for larger smoothing scales as the objects have a smaller angular scale at higher redshifts.

Ideally, the smoothing scale should be matched to the redshift of an object such that an appropriate scale can be chosen that smooths over the entire object but not over a larger area than necessary, as having either a too-small or too-large smoothing scale will decrease the SNR of the object. As the redshift of an object is not known beforehand, it is not clear what the optimal smoothing scale is to detect an object until it has already been detected and identified with an optical counterpart, and the choice of smoothing scale is a trade-off between detecting more objects at higher redshift against lower completeness at lower redshift. We

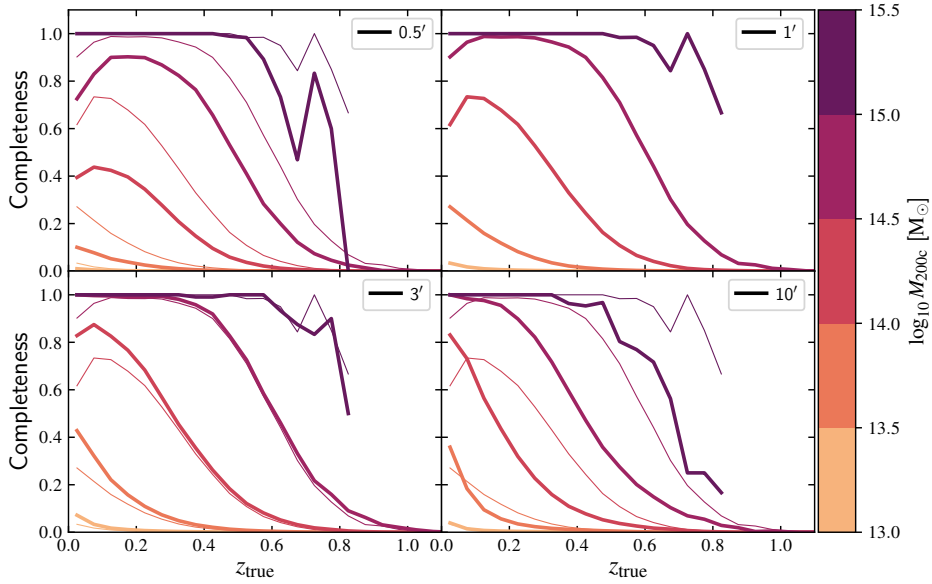


Figure 3.10: Completeness, i.e. the fraction of haloes within a halo mass bin matched to an SNR > 5 WL peak for the four different smoothing scales, as indicated in the upper right of each panel. The results for 1 arcmin smoothing are repeated in all panels as thin curves. For small smoothing scales, some massive haloes at low redshift that span an angular size larger than is smoothed over will be missed as the smoothing scale is too small to achieve the optimal SNR. At the same time, increasing the smoothing scale will result in lower completeness at high redshift as those objects subtend smaller sizes, and increasing the smoothing will decrease the SNR of those objects. Combining smoothing scales may thus improve the amount of information that is extracted.

chose a 1 ′ smoothing scale as the fiducial scale to keep the results consistent with Broxterman et al. (2024). However, we already noted that multiple smoothing scales may be combined to probe information from different scales.

3.B Noise peak distribution

In this Appendix, we compare the number density of WL peaks for our *Euclid*-like sample to a distribution generated only by noise. The noise field is generated as described in Section 3.3.2. In the regular procedure, we add the noise to the WL convergence field and then smooth the map. The number density of WL peaks selected from this map is indicated by the solid curve in Fig. 3.11. Now, we also show the peak distribution that is obtained if we directly determine the peaks

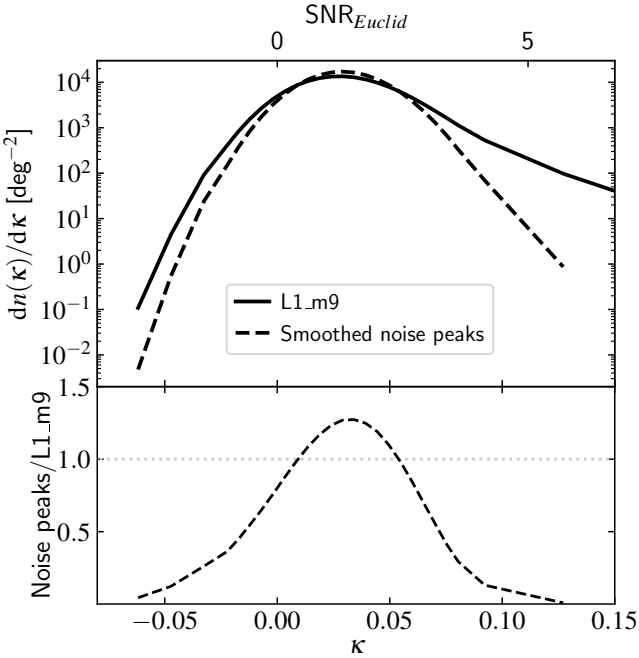


Figure 3.11: Top: number density of WL peaks for the WL convergence map that includes noise and smoothing (solid curve) and of only the smoothed noise (dashed curve). Bottom: ratio of the smoothed noise peak distribution to L1_m9. The smoothed noise peak distribution dominates the signal for $0 \lesssim \kappa \lesssim 0.07$, and lower and higher valued peaks generally correspond to physical objects.

from the smoothed noise map ('Smoothed noise peaks'), which do not contain any cosmological information. The Smoothed noise peak distribution is indicated by the dashed curve. Whereas the noise is a normal distribution centred on $\kappa = 0$, the Smoothed noise peak distribution, consisting of the local maxima of the smoothed noise map, has a positive mean and an asymmetric shape.

In Fig. 3.2, we illustrated that the Noise peak distribution peaks at $\kappa \approx 0.04$. Fig. 3.11 illustrates that this is also the WL convergence value at which the distribution from the applied noise field peaks. As expected, for larger κ values, the distribution of Smoothed noise peak decreases, and the majority of the L1_m9 peaks are no longer caused by the noise, as also already shown in Fig. 3.2. Similarly, the lowest valued WL peaks, with $\kappa \lessapprox 0$, are also increasingly less likely to be caused by the random noise. These peaks, therefore, correspond to the local maxima in void regions in the simulations.

3.C Halo–peak matching characteristics

In this Appendix, we study the characteristics of the peak-halo matching algorithm in more detail. In Fig. 3.12, we show the same fields as in Fig. 3.3, but we split the peaks into the three peak categories we considered in this work. The solid yellow, cyan, and green circles indicate the Single shell, Noise, and Multiple shells peaks, respectively. We also show all $M_{200c} > 10^{13.5}$ $M_\odot$ haloes from the FLAMINGO halo light-cone as dashed-dotted white circles and the matched haloes as dotted red circles.

Fig. 3.12 shows that Noise and Multiple shells peaks may still correspond to massive overdense objects. For example, a Noise peak is indicated by the solid cyan circle in the middle left of the figure. Similarly, the solid green circle directly above the legend shows that this Multiple shells peak can also be assigned to a halo at the same angular position. When comparing the redshift distribution of the peaks, we have also assigned a redshift to the Noise peaks and Multiple shells peaks. In any observational inference, we will not know which peaks are Noise peaks or Multiple shells peaks, but can only assign a redshift based on the galaxies at the position of the peak. The comparison here shows that this is a valid approach, as there are still haloes assigned that contribute significantly.

In orange, we show the WL peaks assigned to this shell with an SNR of $3-5$. These are lower values than the standard range we consider in this work. The figure shows that most of these peaks still clearly correspond to an overdense structure in the field, as indicated by the white dashed-dotted circles. Lowering the SNR threshold would allow us to include these matches in samples we consider and possibly get tighter constraints, but this comes at the cost of decreasing the purity of the sample, as shown in Fig. 3.2. As a result, the interpretation will be more difficult, as a superposition of objects or noise will cause more WL peaks, and interpreting them each as arising from a single halo along the line of sight will no longer be justified. This is also clear from the fact that some of these peaks do not correspond to a massive overdense structure. However, we stress that our chosen SNR cut-off may not be optimal. Additional work should be done to explore the optimal SNR cut-off value to balance the trade-off between cosmological information and purity.

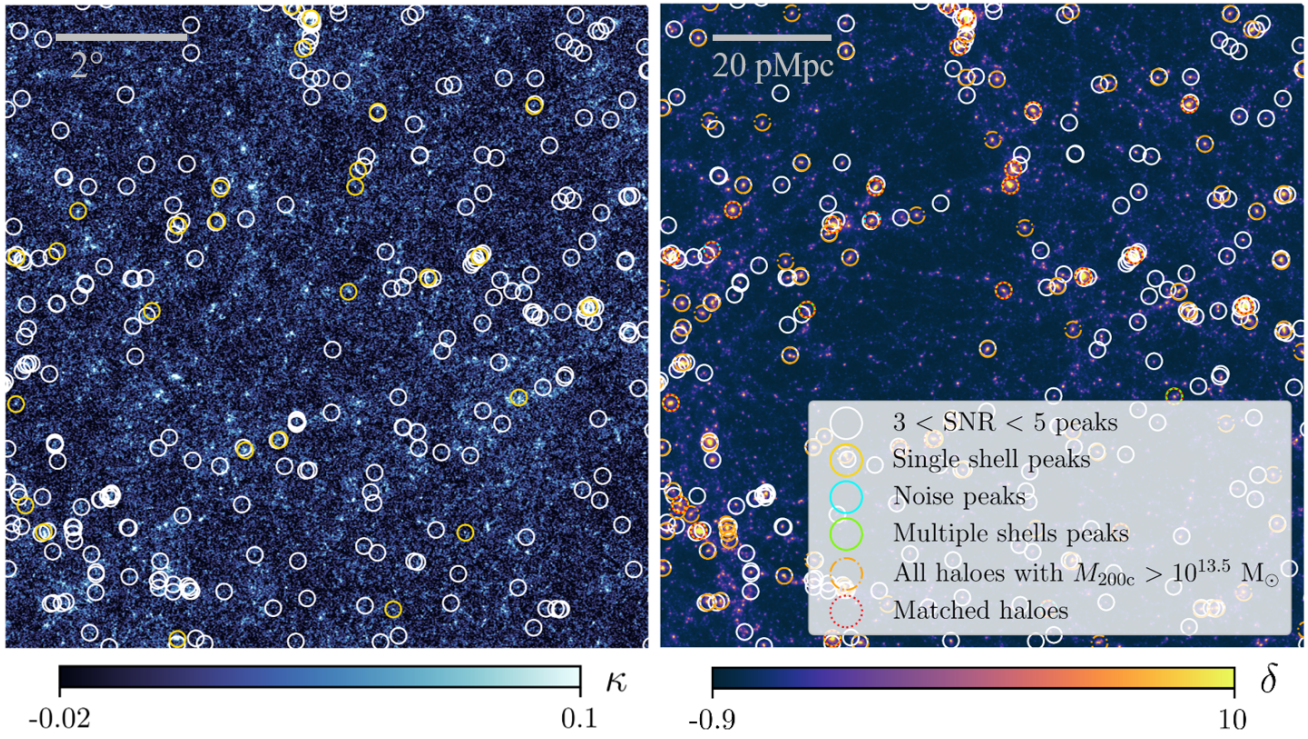


Figure 3.12: Same as Fig. 3.3 and now also showing the different peak categories to which the peaks belong. The Single shell peaks (solid yellow), Noise peaks (solid cyan), Multiple shells peaks (solid green), peaks assigned to this shell with SNR = 3 – 5 (solid white), all $\log_{10} M_{200c}[M_\odot] > 13.5$ haloes (dashed-dotted orange) and haloes that have been matched to an SNR > 5 peak (dotted red). Most Noise and Multiple shells peaks are still visibly well matched to a halo at their position. The apparent overdensities that do not generate an SNR > 5 peak produce peaks with an SNR value lower than this cut.

4 Cosmological constraining power of the redshifts, heights, and angular clustering of weak gravitational lensing peaks

Authors: **Jeger C. Broxterman**, Matthieu Schaller, Ian G. McCarthy, Willem Elbers, John Helly, Henk Hoekstra, Konrad Kuijken, Jaime Salcido, Joop Schaye, Naomi Schutte and Elena Sellentin

Published in Monthly Notices of the Royal Astronomical Society, 553, stag1905 (2026)

Abstract

Weak gravitational lensing (WL) peaks probe non-Gaussian information of the large-scale distribution of matter that is not captured by two-point lensing statistics. We study the cosmological potential of the height distribution, redshift distribution, and angular clustering of high-valued WL peaks using a Bayesian inference approach that mimics a *Euclid* analysis. We use a forthcoming dark-matter-only hypercube, which varies cosmology in a ten-dimensional space, including evolving dark energy, neutrino mass, decaying dark matter, and the running of the scalar spectral index. We find the individual WL peak statistics to be complementary, as the redshift distribution best constrains the matter density, Ω_m, and the dark energy equation-of-state parameters, w_0, and w_a; the height distribution and angular clustering are most sensitive to the amplitude of the primordial power spectrum, $\ln(10^{10}A_\mathrm{s})$; while combining the three statistics allows us to also probe the baryon density, Ω_b, and the Hubble parameter, h. A comparison to the shear two-point correlation function demonstrates that WL peaks alone outperform the commonly used statistics, while even tighter constraints are obtained when combining all. Considering the 2-dimensional $\Omega_\mathrm{m} - \ln(10^{10}A_\mathrm{s})$ and $w_0 - w_a$ parameter planes, we find that the redshift distribution outperforms the angular clustering and peak-height distributions. We study the impact of the smoothing scale and find that, typically, the smallest scales yield the best results and the figure of merit improves by a factor ≈ 2 when combining multiple scales.

4.1 Introduction

The concordance model of cosmology, ΛCDM, is able to describe a plethora of observations with only a very limited number of free parameters and main model ingredients, namely dark energy (DE, Λ if its energy density is constant with cosmic time), cold dark matter (CDM), baryons, and inflation (for a recent review, see e.g. Lahav & Liddle, 2024). However, as current surveys become more precise, hints of deviations from this model are emerging. Specifically, recent Dark Energy Spectroscopic Instrument (DESI) results, which constrain cosmology through baryonic acoustic oscillations (BAO) and the clustering of galaxies, quasars, and the Lyman-alpha forest, provide hints for evolving DE (Adame et al., 2025; Abdul Karim et al., 2025).

Similarly, in its recent Year 6 inference, the Dark Energy Survey (DES), following the Chevallier-Polarski-Linder (CPL) $w_0 - w_a$ parametrisation (Chevallier & Polarski, 2001; Linder, 2003), finds that the best-fitting equation of state of dark energy slightly deviates from the ΛCDM expectation. Although their best-fit measurement from the combination of weak gravitational lensing (WL) and galaxy clustering (3×2pt) is still consistent with no evolving dark energy, when they add their BAO and supernova (SNe) measurements, they find a 2.2σ deviation from a cosmological constant (DES Collaboration et al., 2026b).

One of the main cosmological probes for studying the time evolution of DE is cosmic shear, the coherent distortion of the images of background galaxies caused by gravitational lensing by the large-scale structure of the Universe (see e.g. Hoekstra & Jain, 2008). As the stage III lensing surveys, the Kilo-degree Survey (KiDS; de Jong et al., 2013), DES (The Dark Energy Survey Collaboration, 2005), and Hyper Suprime-Cam (HSC; Miyazaki et al., 2012), are in the final stages of their (extended) cosmological analyses (e.g. Reischke et al., 2026), the next generation is already underway as *Euclid* (Euclid Collaboration: Mellier et al., 2025) is carrying out its first cosmological inference, *Rubin*'s Legacy Survey of Space and Time has started (LSST Science Collaboration et al., 2009), and *Roman* has launched (Spergel et al., 2015). Combined, these instruments are expected to exhaust the sky best suited to cosmic shear, while combining deep observations with high-quality imaging.

Stage III surveys have consistently measured lower $S_8 \equiv \sigma_8(\Omega_\mathrm{m}/0.3)^{1/2}$ values than inferred by *Planck*. Although the extent of these differences is smaller than previously reported (for a recent review, see Pantos & Perivolaropoulos, 2026), next-generation surveys are needed to conclusively address these discrepancies and identify their origins. *Euclid* alone is expected to provide per cent-level constraints on the equation-of-state parameters of DE (Euclid Collaboration: Cañas-Herrera et al., 2025). Comparing the results of these different probes and surveys should help us understand the minor discrepancies currently observed. Extracting all cosmological information from the observed data will be key in resolving

these differences.

The common statistics used by cosmic shear surveys to constrain cosmology are 2-point projections of the matter power spectrum (see e.g. Wright et al., 2025; DES Collaboration et al., 2026a), which quantify the variance of the matter density contrast field as a function of scale. However, more precise constraints may be obtained by including additional observables that probe non-Gaussian information not captured by traditional two-point measurements (e.g. Euclid Collaboration: Ajani et al., 2023; Euclid Collaboration: Vinciguerra et al., 2026). Examples of these statistics are the WL scattering transform (e.g. Cheng et al., 2020), Minkowski functionals (e.g. Marques et al., 2019), Betti numbers (e.g. Feldbrugge et al., 2019), wavelet phase harmonics (Allys et al., 2020), persistent homology (e.g. Heydenreich et al., 2021), or three-point statistics (e.g. Linke et al., 2026).

This paper focuses on another higher-order weak lensing statistic, WL peaks, the local maxima in WL convergence fields (e.g. Jain & Van Waerbeke, 2000; Wang et al., 2009), which typically correspond to the most massive structures in the Universe that have experienced the strongest non-linear collapse (Hamana et al., 2004; Yang et al., 2011). It has been shown that these peaks contain cosmological information in their height distribution (Dietrich & Hartlap, 2010; Kratochvil et al., 2010; Broxterman et al., 2024), steepness (Li et al., 2023), angular clustering (Marian et al., 2013; Davies et al., 2019), and redshift distribution (Broxterman et al., 2025). At the same time, the measurement is only first-order in the observed shear and could therefore respond differently to systematics than two-point statistics. Combined, the different WL peak properties could probe information similar to that from traditional galaxy cluster cosmological inferences, as they are sensitive to the mass and location of massive structures. One major advantage is that the peaks are WL- and thus mass-selected and therefore do not suffer from selection effects that require assumptions about the dynamical state of massive structures, or the form and scatter of mass-to-observable scaling relations, which typically limit X-ray- or thermal-Sunyaev Zel'dovich-selected cluster abundance inferences (see e.g. Kugel et al., 2024; Marini et al., 2024; Kugel et al., 2025).

In a cosmological forecast, Zürcher et al. (2021) showed that for a stage-III-like WL survey, including the peak height distribution can already increase the figure of merit (FoM) in the $\Omega_\mathrm{m}-\sigma_8$ plane by a factor of 4 compared to the angular power spectrum alone. Similarly, Davies et al. (2022) find that the angular clustering of WL peaks is significantly more sensitive to the cosmological parameters h and w_0 than the peak number density distribution, and when the two properties are combined, the constraining power on $\Omega_\mathrm{m}, S_8, h$, and w_0 more than doubles. In this paper, we aim to extend this work by conducting a cosmological forecast for high-signal-to-noise ratio (SNR) WL peaks for a Stage IV WL cosmological inference, combining different WL peak properties to study their complemen-

tarity. We do this by using a new set of simulations in a ten-dimensional (10D) cosmological parameter space that includes evolving dark energy, decaying dark matter (DDM), and a running scalar spectral index in (700 cMpc)^3 volumes (McCarthy et al., in prep). We include the number densities, redshift distribution, and angular clustering of high-valued WL peaks.

In Section 4.2, we introduce the relevant WL theory to infer WL signals from a general matter distribution. In Section 4.3, we introduce the hypercube simulations. In Section 4.4, we demonstrate how WL signals are inferred from the simulation, which peak signals are selected, and present the details of the Bayesian likelihood forecast. The results and discussion are given in Section 4.5, and the final conclusions in Section 4.6.

4

4.2 Theory

In this section, we summarise the main equations to estimate the WL signal from a matter density distribution. For extended reviews of WL and cosmic shear, see Bartelmann & Schneider (2001) and Kilbinger (2015). We assume a spatially flat universe such that the comoving angular diameter distance, $f_K(\chi)$, reduces to the comoving distance, χ. In WL, photon deflections are described via the lensing equation, which relates the unlensed angular position, $\boldsymbol{\beta}$, to the observed angular position $\boldsymbol{\theta}$ as

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \nabla_{\boldsymbol{\theta}}\,\psi(\boldsymbol{\theta}, \chi), \tag{4.1}$$

where ψ is the lensing potential. We assume the Born approximation, i.e., we evaluate the lensing quantities on unperturbed photon paths, such that ψ is given by

$$\psi(\boldsymbol{\theta}, \chi) = \frac{2}{c^2}\int_0^{\chi} \mathrm{d}\chi'\,\frac{\chi - \chi'}{\chi\chi'}\,\phi(\boldsymbol{\theta}\chi', \chi'), \tag{4.2}$$

where c is the speed of light in vacuum and ϕ is the Newtonian gravitational potential. Conventionally, the lens equation is linearised to first order in the lensing potential, and the effect is quantified in terms of the magnification matrix, A, as

$$\frac{\partial\beta_i}{\partial\theta_j} \equiv A_{ij} = \delta_{ij} - \partial_i\partial_j\psi, \tag{4.3}$$

where the magnification matrix is decomposed as

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{4.4}$$

where $\gamma = \gamma_1 + i\gamma_2$ is the WL shear and κ is the WL convergence, which is used here to quantify the cosmic shear strength. Equations 4.2–4.4 can be combined

to express the WL convergence as

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{1}{c^2} \int_0^{\chi} \mathrm{d}\chi' \frac{\chi - \chi'}{\chi' \chi} \nabla^2_{\boldsymbol{\theta}} \phi(\chi' \boldsymbol{\theta}, \chi'). \tag{4.5}$$

We use the 3D Poisson equation to estimate the deflection strength

$$\nabla^2 \phi = 4\pi \overline{\rho_{\mathrm{m}}} a^2 G \delta \approx \chi^{-2} \nabla^2_{\boldsymbol{\theta}} \phi, \tag{4.6}$$

where G is the gravitational constant, $\overline{\rho_{\mathrm{m}}}$ is the mean matter density, a the scale factor, and δ the matter density contrast, and, in the final step, we assume that the contributions along the line of sight cancel, i.e. the Limber approximation (Limber, 1953). To account for a population of source galaxies, we integrate Eq. 4.5 as

$$\kappa(\boldsymbol{\theta}) = \int_0^{z_{\mathrm{hor}}} \mathrm{d}z \, n_{\mathrm{s}}(z) \, \kappa[\boldsymbol{\theta}, \chi(z)], \tag{4.7}$$

where the normalised source redshift distribution, n_{s}, extends to the edge of the survey, $z_{\mathrm{hor}} = 3$ in our case. We assume a *Euclid*-like source redshift distribution given by (Euclid Collaboration: Blanchard et al., 2020)

$$n(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right], \tag{4.8}$$

with $z_0 = 0.9/\sqrt{2}$, such that the median redshift is 0.9.

Our analysis ignores several higher-order WL effects, contaminants, and the possibly non-negligible impact of approximations, such as the observable being shear rather than reduced shear (e.g. Seitz & Schneider, 1995; Bradač et al., 2005), source clustering (e.g. Gatti et al., 2024), intrinsic alignments (IA; e.g. Lee et al., 2026; Vedder et al., 2026), and the Born approximation (e.g. Lu & Haiman, 2021). We also do not marginalise over nuisance parameters such as magnification bias (e.g. Liu et al., 2014) or shifts in the mean or shape of the redshift distribution (e.g. Hildebrandt et al., 2020). However, we assume these effects do not affect the relative differences we study in our idealised inference. Nevertheless, we stress that any actual inference should consider these effects in greater detail. While a quantitative assessment of these systematics requires a more elaborate study, we can draw some qualitative expectations from their known dependencies. Baryonic feedback suppresses the amplitude and abundance of the highest-convergence WL peaks, since these are sourced by the most massive collapsed structures. Because the redshift and height distributions in this work are built from the same underlying peak population, they are likely to respond to baryonic suppression in a correlated way, as a systematic reduction in the number and amplitude of the most massive haloes would suppress both statistics coherently. However, as shown in Broxterman et al. (2025), baryonic feedback primarily changes the amplitude of the redshift distribution and not its shape, which

is affected by cosmology. Similarly, angular clustering at fixed WL peak number density is primarily sensitive to the spatial distribution of peaks rather than their amplitude and is thus naively expected to be less sensitive to baryonic feedback. Because we work directly with the convergence field from mass maps, IA, which arises from correlated intrinsic source-galaxy shapes, does not enter our analysis. For the height distribution, Zhang et al. (2022) show that IA impacts the height distribution in a non-negligible way. Given that our redshift assignment distribution is based on properties from the observed KiDS bright sample (as explained in more detail in Section 4.2), which includes IA, the redshift distribution is expected to be less impacted by this systematic.

4.3 Simulations

In our analysis, we use a new set of simulations, which will be presented in detail in McCarthy et al. (in preparation). We use two sets of 50 spatially flat simulations in 10D and 9D cosmological parameter spaces. The initial conditions have identical phases, and the amplitudes of the large-scale modes are fixed to the mean variance. The parameters that are varied are detailed in Table 4.1, and include the present-day matter density parameter in the limit of no DDM, $\Omega_{\rm m,noDDM}$, i.e. the value that $\Omega_{\rm m}$ would have been if the simulation had no DDM, the cosmic baryon fraction given by the ratio of the redshift zero baryon density to matter density parameters, $f_{\rm b}$, the Hubble parameter, h, the scalar spectral index, $n_{\rm s}$, the amplitude $\ln(10^{10}A_{\rm s})$, and the running of the scalar spectral index, $\alpha_{\rm s} \equiv {\rm d}n_{\rm s}/{\rm d}\ln k$, of the primordial power spectrum, the dark energy equation-of-state parameters following the CPL parametrization, w_0 and w_a, the sum of 3 massive neutrino species, $\sum m_\nu c^2\,[{\rm eV}]$, and the decay rate of decaying cold dark matter, $\Gamma_{\rm dcdm}[{\rm km/s/Mpc}]$ following the model introduced in Elbers et al. (2025b). In the DDM extension, a fraction of the cold dark matter is treated as an unstable species with mean lifetime τ, decaying into a relativistic dark radiation component. $\Gamma_{\rm dcdm} = \tau^{-1}$ is the decay rate. The decay is implemented by adjusting the particle masses in proportion to the fraction of dark matter that has decayed.

The sampling within the 10D hypercube follows a Latin hypercube design, which is approximately uniform in the marginal distribution of each parameter individually, with the exception of w_0 and w_a. w_0 and w_a are not sampled uniformly independently, as large parts of the $w_0 - w_a$ parameter plane are ruled out by various observations. Instead, they are sampled along a degeneracy direction that is approximately parallel to the geometric degeneracy from the cosmic microwave background (CMB), DESI BAO and SNe (Abdul Karim et al., 2025). The displacement along the degeneracy direction is varied by $\pm\,0.5$, which brackets all current constraints from the individual probes. The 9D hypercube can be viewed as a projection of the 10D hypercube onto the subspace where the DDM rate is fixed to zero, while the remaining nine parameter values are identical in both hy-

Table 4.1: Varied parameters and their ranges in the hypercube. From top to bottom, the present-day total matter density parameter in the limit of no DDM, $\Omega_{\mathrm{m,noDDM}}$, the baryon fraction, f_{b}; the dimensionless Hubble parameter, h; the scalar spectral index, n_{s}, and the amplitude, $\ln(10^{10}A_{\mathrm{s}})$, of the primordial power spectrum; the CPL equation-of-state parameters of dark energy w_0 and w_a; the sum of neutrino masses, $\sum m_\nu c^2$ [eV]; the running of the spectral index, α_{s} of the primordial power spectrum; and the DDM decay rate, Γ_{dcdm} [km/s/Mpc].

Parameter	Hypercube range
$\Omega_{\mathrm{m,noDDM}}$	$[0.2, 0.4]$
f_{b}	$[0.14, 0.175]$
h	$[0.6, 0.8]$
n_{s}	$[0.9, 1.0]$
$\ln(10^{10}A_{\mathrm{s}})$	$[2.965, 3.115]$
w_0	[-1.2, -0.3]
w_a	[-3.7(w_0+1.0), -3.7(w_0+1.0)+0.5]
$\sum m_\nu c^2$ [eV]	$[0, 0.45]$
α_{s}	[-0.03, 0.03]
Γ_{dcdm} [km/s/Mpc]	$[0, 20]$

percubes. Combining the two hypercubes results in a joint 100-node training set in which the $\Gamma_{\mathrm{dcdm}} = 0$ hyperplane is sampled more densely than any other value of Γ_{dcdm}. In our forecasts, we always jointly use these two hypercube sets to train the emulators.

Each hypercube node corresponds to a single simulation realisation, with no repeated realisations at fixed cosmology to sample cosmic variance or shape noise. While we find this is sufficient to obtain unbiased constraints in our simulation forecast using mode fixing and inferring the cosmology from a simulation run within the same framework, it may not be sufficient for an inference on real data.

The hypercube simulations were run using the same set-up as the original FLAMINGO simulations (Schaye et al., 2023; Kugel et al., 2023), only in smaller volumes. The simulations were run using SWIFT (Schaller et al., 2024) and each contains 1260^3 DM particles and 700^3 massive neutrino particles in a box with a side length of 700 cMpc, which results in a typical initial DM particle mass of a few $\times 10^9$ $\mathrm{M}_\odot$, depending on the exact values of the cosmological parameters. The primordial spectrum, depending on $\ln(10^{10}A_{\mathrm{s}})$, n_{s}, and α_{s} together with the transfer functions computed with CLASS (Lesgourgues, 2011), sets the linear matter power spectrum used to generate the initial conditions using MONOFONIC (Hahn et al., 2020; Michaux et al., 2021). Each simulation has a virtual observer with a spherical full-sky light-cone from which we use the matter shells.

We ignore the impact of baryonic physics, i.e., the redistribution of matter due to black hole feedback, supernova explosions, or other physical processes pertaining to baryons that impact the distribution of matter on (mildly) non-linear scales, even though this is one of the dominant systematics in WL inferences (see e.g., Semboloni et al., 2011, 2013). In Broxterman et al. (2024, 2025), we found that the WL peak height distribution and redshift distribution depend differently on cosmology and baryonic physics, and combining the two statistics helps to simultaneously constrain cosmology and calibrate baryonic physics. In this forecast, we also include the angular clustering of these peaks, which will depend in its own way on baryonic physics, and therefore may improve the baryonic feedback calibration. At the same time, we stress that this should be tested using hydrodynamical simulations and that we do not expect this to affect the relative differences between our results for the constraining power of the WL peak properties.

4.4 Methods

In this section, we present the pipeline for deriving full-sky WL convergence maps from maps of the matter density, and the subsequent measurements of the WL peak statistics and WL shear two-point correlation functions from these maps. Afterwards, we present the inference pipeline consisting of the emulator, likelihood, and covariance matrix, which will be used to infer cosmological constraints from an unseen simulation run.

The construction of the WL convergence map closely follows the method described in Broxterman et al. (2025). We use the concentric Healpix (Górski et al., 2005) light-cones at $N_{\rm side} = 8192$ that are centred around a virtual observer at the same position in each simulation. The shells have a thickness of $\Delta z = 0.05$. Full-sky WL convergence maps are constructed from spherical matter shells by discretising Eq. 4.7 as

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{4\pi G}{c^2} \sum_i [1 + z(\chi_i)]^{-2} \Delta\chi_i \, \delta_i(\boldsymbol{\theta}) \Omega_{\rm m}(\chi_i) \rho_{\rm crit}(\chi_i) W(\chi_i), \tag{4.9}$$

$\rho_{\rm crit}$ is the critical energy density of the universe, $\Omega_{\rm m}(\chi_i) = \rho_{\rm m}(\chi_i)/\rho_{\rm crit}(\chi_i)$ is the value of the matter density parameter at χ_i, whose time evolution is different from $w_0 w_a$CDM as the simulations include DDM, and $\Delta\chi_i$ is the comoving thickness of light-cone shell i. We evaluate the matter density contrast, δ, directly from the total matter shells. Each time the light-cone diameter exceeds an integer multiple of the box size, we apply a random rotation to all shells within the next integer multiple using the same rotation angles, thereby avoiding spurious discontinuities at the box boundary while still suppressing repeated large-scale

structure along the line of sight. The WL kernel is given by

$$W(\chi_i) = \chi_i \sum_{j=i}^{N_{\rm shells}} n_{\rm s}(\chi_j)\left(1 - \frac{\chi_i}{\chi_j}\right)\Delta\chi_j, \tag{4.10}$$

with $n_{\rm s}(\chi){\rm d}\chi = n_{\rm s}(z){\rm d}z$ and which runs until $N_{\rm shells} = 60$. We mimic shape noise by assigning a randomly drawn value from a Gaussian with mean μ and standard deviation σ

$$\mathcal{N}\left\{\mu = 0, \sigma = \frac{\sigma_\epsilon}{\sqrt{2 n_{\rm gal} A_{\rm pix}}}\right\}, \tag{4.11}$$

where $A_{\rm pix}$ is the area of one pixel (0.18 arcmin^2 for $N_{\rm side} = 8192$), $n_{\rm gal}$ is the source galaxy number density, and σ_ϵ is the root mean square total intrinsic ellipticity of source galaxies. To mimic a *Euclid* survey, we take $n_{\rm gal} = 30$ arcmin^{-2}, and $\sigma_\epsilon = 0.26\sqrt{2}$ (Euclid Collaboration: Mellier et al., 2025).

4.4.1 Smoothing

We explore the impact and complementarity of combining different smoothing scales. In our fiducial setup, similarly to Broxterman et al. (2024, 2025), we smooth the integrated WL convergence map using a Gaussian kernel with a full width at half maximum of 1 arcmin, corresponding to 2.3 pixels. This was identified by Liu et al. (2015) as close to the optimal smoothing scale for their cosmological inference from CFHTLenS data, based solely on the WL peak number density. However, they already highlight that combining multiple smoothing scales may improve constraining power, as different smoothing scales correspond to different spatial scales. As we, instead, mimic a *Euclid*-like analysis, with a different resolving power, and we combine multiple peak statistics, we reevaluate the impact of varying the smoothing scale on the constraining power. When selections are made based on the SNR, we recompute the noise standard deviation from the smoothed, randomly generated noise realisation for each smoothing scale. In Section 4.5.5, we compare the fiducial results at 1 arcmin smoothing to those obtained with a smoothing scale that is a factor of 2 larger or smaller, or when combining the three smoothing scales.

4.4.2 Peak statistics

We use three different WL peak statistics in our forecast, the redshift distribution, ${\rm d}n(z)/{\rm d}z$; height distribution, ${\rm d}n(\kappa)/{\rm d}\kappa$; and angular clustering, $\omega(\theta)$. More details on the binning are given in Appendix 4.A. The peaks are selected from the integrated WL convergence maps after noise and smoothing have been applied. Peaks are defined as pixels with values greater than those of their eight nearest neighbours, as commonly used in WL peak studies (e.g., Weiss et al., 2019;

Grandón et al., 2024). For each peak, we determine its WL convergence value, subsequently referred to as the 'peak height'. For SNR > 3 peaks, we identify the primary redshift that generates each peak, following the procedure of Broxterman et al. (2025). In the simulations, using the mass maps, the redshift of the overdensity that contributes most to the WL convergence at each WL peak, i.e. the dominant contribution to Eq. 4.9, is assigned to that peak.

The redshifts are distorted, mimicking the observational uncertainty in photometric redshift estimation for massive galaxy clusters. As the high-value WL peaks generally correspond to massive clusters, the redshifts of these objects are expected to be measured with much greater accuracy than those of the average *Euclid* galaxy. Whereas the *Euclid* requirement on the precision of the photo-z estimates is $\sigma_z < 0.05(1+z)$ with catastrophic failures rate of less than 10 per cent (Euclid Collaboration: Mellier et al., 2025), these objects are expected to be measured more accurately using regular photo-z estimation procedures or red-sequence cluster finders (see e.g. Rykoff et al., 2014; Chiu et al., 2024; Chen et al., 2025). Based on the properties of the KiDS-1000 bright sample (Bilicki et al., 2021), which were estimated using a similar number of photometric bands to those that will be available for *Euclid*, the simulation redshifts are distorted by, for each peak, drawing a random redshift from a normal distribution centred on the true redshift, $z_{\rm true}$, and with a standard deviation of $\sigma_z = 0.02(1+z_{\rm true})$, without any catastrophic outliers. The WL peak redshift distribution thus aims to exploit these precise redshift measurements of these massive objects for cosmological purposes.

As a third statistic, for the SNR > 3 peaks, we estimate the WL peak angular two-point correlation function. We use the Landy-Szalay estimator (Landy & Szalay, 1993)

$$\omega(\theta) = \frac{\mathrm{DD}(\theta) - 2\mathrm{DR}(\theta) + \mathrm{RR}(\theta)}{\mathrm{RR}(\theta)}, \tag{4.12}$$

to estimate the clustering signal. Here, θ is the angular separation between two WL peaks, and DD, RR, and DR represent the number of pairs of peaks separated by θ in the data sample (DD), a uniformly randomly generated sample (RR), and their cross-catalogue (DR). In practice, we use the package Corrfunc (Sinha & Garrison, 2020) to estimate the angular correlation function. The random set is chosen to be 8 times larger than the data catalogue in order to suppress Poisson noise originating from the random catalogue.

Although we probe the angular clustering and redshift distribution of the high-SNR peaks, these could, in theory, also be combined into a single statistic measuring the 3D clustering. This is, however, beyond the scope of this work.

4.4.3 Two-point correlation functions

To compare our forecast with those similar to a typical WL inference, we additionally estimate the WL angular two-point correlation functions (2PCFs), $\xi_{+/-}$. We do so by computing the WL convergence angular correlation function, $\mathcal{C}^{\kappa\kappa}$, from the full-sky WL convergence maps, which in the absence of B-modes relates to $\xi_{+/-}$ through a Hankel transform as (Schneider et al., 2002)

$$\xi_{+/-}(\theta) = \int_0^\infty \frac{\mathrm{d}\ell\,\ell}{2\pi} \mathcal{C}^{\kappa\kappa}(\ell) J_{0/4}(\ell\theta), \tag{4.13}$$

where $J_{0/4}$ are Bessel functions of the first kind, and ℓ is the multipole moment. We measure the 2PCFs in 9 logarithmic bins between 0.1 and 3 deg, which roughly corresponds to the ranges used by current cosmic shear inferences (see e.g. DES Collaboration et al., 2026a), and is the same angular range as used for the WL peak angular clustering.

4.4.4 Inference forecast

In this section, we provide details of the forecasting pipeline we use to generate our forecasts. As explained in Section 4.3, the WL peak statistics are extracted through forward modelling the hypercube simulations. We then train an emulator to predict the peak statistics as a function of the 10 cosmological parameters that are varied in the hypercube (Section 4.4.4). Finally, we estimate a covariance matrix from additional simulations (Section 4.4.4) and perform a Bayesian likelihood inference (Section 4.4.4).

Emulation

To predict the WL peak statistics across the 10D parameter space spanned by the hypercube, we use Gaussian process (GP) emulators trained on statistics computed from the 2×50 simulations. All emulators are implemented with the GaussianProcessRegressor class from SCIKIT-LEARN (Pedregosa et al., 2011). We train separate emulators for the different statistics and smoothing scales, and concatenate them to produce the combined forecasts. The details of the emulator construction as well as a leave-one-out test are presented in Appendix 4.A. In general, we find sub-per-cent-level relative emulator accuracy across the ranges on which the statistics are used, which is sufficient for our idealised forecast.

Covariance matrix

To estimate the covariance, we run an additional suite of flat ΛCDM simulations. This set uses the same framework as the hypercube, but assumes the fiducial D3A cosmology of Schaye et al. (2023), i.e. a flat ΛCDM ($w_0 = -1, w_a = 0$) without DDM ($\Gamma_{\mathrm{dcdm}} = 0$ km s^{-1} Mpc^{-1}), no running of the scalar spectral index ($\alpha_\mathrm{s} =$

0), only a single massive neutrino species with $m_\nu c^2 = 0.06$ eV, and the same cosmology as the reference cosmology we are trying to infer. We assume that the covariance over the entire cosmological parameter space can be approximated by the covariance generated from these ΛCDM simulations. This assumption is expected to be a subdominant effect for cosmic shear 2pt inferences (see e.g. Kodwani et al., 2019), but has not yet been studied for the non-Gaussian statistics we consider. For each simulation, we use a different random seed to generate the initial conditions with MONOFONIC (Hahn et al., 2020; Michaux et al., 2021), and we do not fix any modes, unlike the simulations that are part of the hypercube. As such, our covariance matrix should reflect the full cosmic variance that is inherent to the inference.

For each forecast and for realisation i, we compute the relevant WL peak statistics, $\mathbf{x}_i$. The sample covariance, $\hat{\mathbf{C}}$, is then computed as

$$\hat{\mathbf{C}} = \frac{1}{N_{\rm cov} - 1} \sum_{i=1}^{N_{\rm cov}} (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^\top , \tag{4.14}$$

where $\bar{\mathbf{x}} = N_{\rm cov}^{-1} \sum_i \mathbf{x}_i$ is the sample mean. Since our analysis mimics a *Euclid*-like survey covering $f_{\rm sky} = 15\,000/41\,253$ of the sky, the covariance is rescaled as

$$\mathbf{C} = \frac{\hat{\mathbf{C}}}{f_{\rm sky}}. \tag{4.15}$$

When multiple statistics or smoothing scales are combined into a joint data vector, the covariance matrix is estimated from the same realisations, thereby capturing all cross-correlations among them. In Appendix 4.B, we show the correlation matrix for our fiducial combined WL peak forecast, i.e. the height distribution, redshift distribution and angular clustering at 1 arcmin smoothing, illustrating that the angular correlation function exhibits strong correlations between different angular bins, whereas the cross-correlations between the different statistics and the correlations between individual bins of the height distribution and the redshift distribution are weaker.

Likelihood

For our forecast, we assume an explicit likelihood. Whereas simulation-based inferences that learn the relationship between cosmological parameters and observables directly could, in theory, provide more precise constraints (Vinciguerra et al., 2026; Williamson et al., 2026), we use an explicit likelihood because it allows us to gain more physical insight into the constraining properties of different statistics. We assume the likelihood follows a multivariate t-distribution to account for the fact that the covariance matrix $\mathbf{C}$ is estimated from $N_{\rm cov} = 80$ simulations rather than being known exactly (Sellentin & Heavens, 2016). The

likelihood is given by

$$\mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) \propto \left\{1 + \frac{[\boldsymbol{x} - M(\boldsymbol{\theta})]^\top \mathbf{C}^{-1} [\boldsymbol{x} - M(\boldsymbol{\theta})]}{N_{\rm cov} - 1}\right\}^{-N_{\rm cov}/2}, \tag{4.16}$$

where the data vector $\mathbf{x}$ consists of (a combination of) the different WL peak statistics, $\mathbf{C}^{-1}$ is the inverse sample covariance matrix, and $M(\boldsymbol{\theta})$ is the emulator prediction. In practice, we evaluate the log-likelihood, which is given by

$$\begin{aligned}\ln \mathcal{L} = &\ln\Gamma\left(\frac{N_{\rm cov}}{2}\right) - \ln\Gamma\left(\frac{N_{\rm cov} - N_{\rm dv}}{2}\right) - \frac{N_{\rm dv}}{2}\ln\left[\pi(N_{\rm cov} - 1)\right] \\ &- \frac{1}{2}\ln|\mathbf{C}| - \frac{N_{\rm cov}}{2}\ln\left(1 + \frac{[\boldsymbol{x} - M(\boldsymbol{\theta})]^\top \mathbf{C}^{-1} [\boldsymbol{x} - M(\boldsymbol{\theta})]}{N_{\rm cov} - 1}\right),\end{aligned} \tag{4.17}$$

where Γ is the gamma function and $N_{\rm dv}$ is the dimension of the data vector $\boldsymbol{x}$. In the analysis, we infer the cosmology of the reference data vector, $\boldsymbol{x}$, for which we run an additional simulation using the hypercube infrastructure but at the fiducial cosmology of the FLAMINGO simulations (Schaye et al., 2023), which is the best-fitting Dark Energy Survey year 3 3×2pt + All Ext. ΛCDM cosmology (D3A; Abbott et al., 2022) with a single massive neutrino species, from which we extract the same statistics through forward modelling. We infer the best-fitting parameters by sampling the log-likelihood with the Markov chain Monte Carlo sampler EMCEE (Foreman-Mackey et al., 2013), and our results are plotted with GETDIST (Lewis, 2025).

For most parameters, we apply flat priors. For the two beyond $w_0 w_a$CDM parameters, $\alpha_{\rm s}$ and $\Gamma_{\rm dcdm}$, we apply, respectively, a Gaussian and half-Gaussian prior centred on the truth with widths 0.02 and 1, to penalise strong deviations from these less common extensions.

We quantify the FoM for a 2D cosmological parameter combination of θ_1 and θ_2 as

$$\text{FoM} = \frac{1}{\sqrt{|C(\theta_1, \theta_2)|}}, \tag{4.18}$$

where C is the 2×2 parameter covariance matrix of θ_1 and θ_2. Similarly, we quantify the figure of bias (FoB) as

$$\text{FoB} = \sqrt{(\theta_{\rm truth} - \theta_{\rm inferred})^\top C(\theta_{\rm inferred})^{-1} (\theta_{\rm truth} - \theta_{\rm inferred})}, \tag{4.19}$$

where $\theta_{\rm truth}$ is the value of the cosmological parameters in the D3A cosmology and $\theta_{\rm inferred}$ is the best-fit emulator prediction.

4.5 Results and Discussion

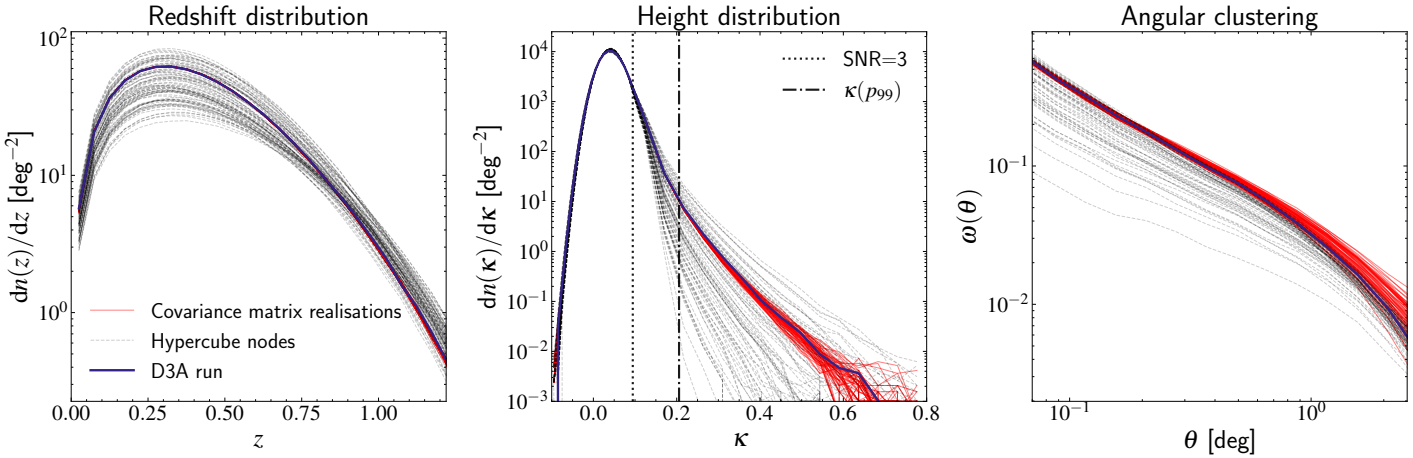


Figure 4.1: WL peak statistics used in the fiducial forecast with 1 arcmin smoothing. Left: redshift distribution of SNR > 3 peaks. Middle: height distribution. Right: angular correlation function of SNR > 3 peaks. The measurements from the hypercube simulations are shown as dashed light-grey curves, the covariance matrix realisations as red lines and the measurements from the additional simulation at the FLAMINGO D3A cosmology, which is inferred in this paper, as a purple curve. The dotted and dashed-dotted lines in the central panel indicate the SNR = 3 threshold used for all statistics and the κ value corresponding to the 99th-percentile peak, which is used as an upper bound for the peak-height distribution.

In Fig. 4.1, we illustrate the measurements of the three different WL peak statistics across the different parts of the forecast pipeline. The left panel shows the redshift distribution of the SNR > 3 peaks, the central panel shows the height distribution, and the right panel shows the angular two-point correlation function of the SNR > 3 peaks. These measurements correspond to our fiducial smoothing scale of 1 arcmin. In each panel, the solid red curves indicate the measurements from the covariance-matrix realisations, the solid purple curve indicates the target distribution measured from the D3A simulation, and the dashed grey curves show the measurements in the 100 hypercube nodes. The hypercube nodes span a wide range of predictions, demonstrating sensitivity across the full 10D parameter space. The central panel also includes the SNR = 3 cut as a vertical dotted line.

For each statistic, the D3A measurement falls within the range of the covariance realisations. The spread of the covariance-matrix realisations relative to the D3A target varies across the three statistics, reflecting the distinct cosmic-variance properties of each observable. For the redshift distribution, the covariance realisations show very little scatter around the D3A measurement across the entire redshift range, reflecting the small sample variance on the shape and amplitude of the redshift distribution of high-SNR WL peaks expected for the forecasted *Euclid* measurements. For the height distribution, the spread of the covariance realisations increases for κ values above the SNR cut. This is expected, as the number density of the peaks drops steeply with increasing κ, and the Poisson noise from the rare high-mass haloes that generate these peaks thus increases. Additionally, for the highest mass objects, the relative location of the (virtual) observer also matters, as it determines the exact peak height via the lensing kernel. As discussed in more detail in Appendix 4.A, we do not use the highest 1 per cent κ peaks during emulator training and inference for the height distribution, as we found the emulator performed better without them, likely due to increased stochasticity. For the angular clustering, the covariance realisations show a scatter of a factor of a few in the signal across the entire angular range, with the relative spread increasing toward larger angular separations.

4.5.1 The impact of cosmology

Before examining the actual cosmological forecast, we first discuss the possible ways in which the changes in cosmology may impact our WL peak forecast. The 10 cosmological parameters varied in the hypercube can affect the WL peak statistics in three distinct physical ways.

The first possibility is through the direct impact on the physical matter density. From Eq. 4.9, κ is directly proportional to $\Omega_{\rm m}(\chi_i)\rho_{\rm crit}(\chi_i)$, i.e. the physical total matter density over time. Within the cosmologies in the hypercube, $\rho_{\rm crit}(z) = 3H(z)^2/(8\pi G)$ and the physical matter density thus scales as $\propto \Omega_{\rm m}h^2$,

illustrating that Ω_m and h directly impact the WL signals through this channel. Additionally, Γ_{dcdm} impacts the total matter density, and therefore directly rescales the amplitude of each peak, albeit with a small magnitude as the DDM rates are small.

The second way in which the WL peaks can be impacted by cosmology is through a change in the growth of structures, which sets the matter density contrast $\delta_i(\boldsymbol{\theta})$ in Eq. 4.9. Structure growth controls the masses of the structures that generate the WL peaks, thereby also controlling the number density and height of the peaks. Parameters that impact the structure growth directly are the amplitude of the primordial power spectrum, $\ln(10^{10}A_s)$, which sets the overall strength of initial density fluctuations, the sum of neutrino masses, $\sum m_\nu c^2$, whose free-streaming suppresses structure growth, but also DE (w_0, w_a), as it may suppress the growth of matter perturbations through its impact on the growth factor.

Finally, the cosmological parameters may impact the WL peak properties through changes in the geometry, i.e. by changing the comoving distance–redshift relation, $\chi(z)$. Therefore, any parameter that alters the expansion history through the Friedmann equation changes the WL kernel (Eq. 4.10) and the conversion between physical separations and angles on the sky, which is relevant for the angular clustering signal.

Although it is not possible to cleanly separate the impact of these different channels on the parameter posteriors, in the next sections, we discuss possibilities for how individual parameters may explain the differences in constraining power we observe.

4.5.2 Cosmological constraints forecast

Next, we forecast the cosmological constraints from our Bayesian analysis for our fiducial set-up for each of the peak statistics individually, and for the forecast that combines the three statistics. In Fig. 4.2, we show the marginalised one-dimensional posterior distributions for all 10 cosmological parameters for each of the three individual statistics and their combination, at the fiducial smoothing scale of 1 arcmin. The forecast using the height distribution, redshift distribution, angular clustering, or the combined approach yields the red, orange, blue, and black posteriors, respectively. The shapes of the priors are indicated by the grey dotted line and grey shaded area. In Appendix 4.C, we show the full corner plot corresponding to forecasts using the 3 WL peak statistics together, the redshift distribution individually, and the angular clustering and height distributions together. This comparison shows the value of just adding the redshift distribution over the other two statistics.

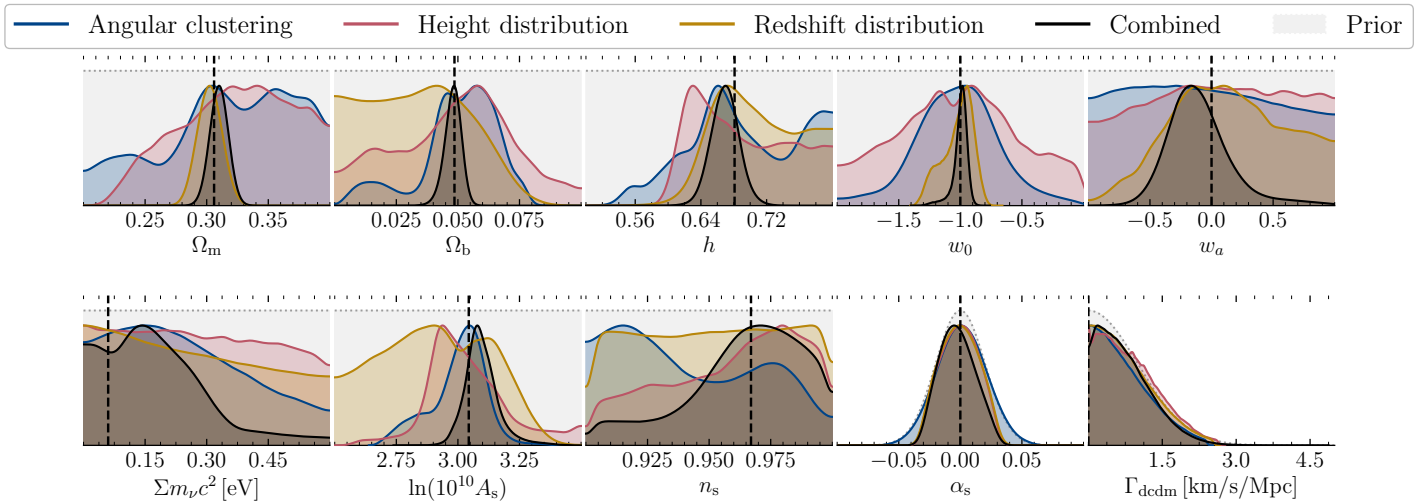


Figure 4.2: One-dimensional posteriors of the fiducial forecast using WL peak angular clustering (blue), height distribution (red), redshift distribution (orange), or the three statistics combined (black) for SNR > 3 WL peaks with 1 arcmin smoothing. The prior is indicated by the dotted grey curve and the shaded grey area, which uniformly spans the parameter range for all panels except the final two. The true values are the D3A cosmology indicated by the dashed black lines. For the combined inference, informative unbiased constraints are recovered for the matter and baryon densities, the Hubble parameter, the DE equation of state, and the amplitude of the primordial power spectrum, whereas the neutrino mass, the running and amplitude of the scalar spectral index and the DDM rate are either poorly constrained or largely recover the prior.

We first assess whether the inference correctly recovers the fiducial D3A cosmology. The combined inference of the three peak statistics yields strong constraints on Ω_m, Ω_b, $\ln(10^{10}A_s)$, w_0, w_a, and h whereas the parameters $\sum m_\nu c^2$, n_s, α_s and Γ_{dcdm} are either weakly constrained or recover the prior. The combined inference (black posteriors) recovers the input parameter values well for those parameters that are well-constrained, as the marginalised posteriors are consistent with the true values (dashed black curves) within the 68 per cent credible intervals, as shown in Fig. 4.2. The average FoB of the maximum a posteriori (MAP) of the parameters that are well constrained, $\Omega_m, \Omega_b, h, w_0, w_a$, and $\ln(10^{10}A_s)$ is 0.59σ. The correspondence demonstrates that the entire pipeline of emulator, likelihood, and covariance matrix together form a self-consistent and proper framework for our cosmological forecasts, and that the added noise does not severely bias the parameter constraints.

Although we expect our idealised inference to recover all parameters well, minor offsets from the truth can be due to (a combination of) noise in the likelihood arising from the finite number of covariance realisations, emulation prediction error, the covariance matrix not varying with cosmology, the added noise components, or because of cosmic variance in the D3A run whose cosmology is inferred. We find some minor offsets from the truth, particularly when a parameter is not well constrained, as is visible in the multi-modal blue angular clustering posteriors for Ω_m, h, and n_s.

Next, we discuss the constraining power on each cosmological parameter and the complementarity between the different WL peak statistics in probing different parts of the cosmological parameter space.

The matter density and amplitude of the primordial power spectrum

The matter density parameter, Ω_m, and the amplitude of the primordial power spectrum, $\ln(10^{10}A_s)$, are among the best-constrained parameters in our forecast, which is expected for Ω_m since it impacts the inference through each channel as discussed before, whereas $\ln(10^{10}A_s)$ only acts directly through the second channel. For the combined inference we find $\Omega_m = 0.310 \pm 0.006$ and $\ln(10^{10}A_s) = 3.09 \pm 0.06$, while the input values were $\Omega_m = 0.306$ and $\ln(10^{10}A_s) = 3.04$. The corresponding best-fitting S_8 is 0.819 ± 0.011, compared to the true value of 0.815. Since we do not marginalise over any systematics, we limit our interpretation to checking that the inferred values are consistent with the truth, rather than drawing conclusions about the absolute constraining power.

The strong constraining power for Ω_m and $\ln(10^{10}A_s)$ is consistent with the results of cosmic shear surveys that typically report their cosmology constraints in terms of the S_8 parameter (e.g. Abbott et al., 2022; Wright et al., 2025), since this is the combination in the Ω_m–σ_8 plane that current cosmic shear surveys are, approximately, most sensitive to. In our forecast, we find that the constraint

on Ω_m is driven by the redshift distribution, whereas the angular clustering and height distribution both contribute significantly to the constraint on $\ln(10^{10}A_s)$.

The sensitivity of the three WL peak statistics to Ω_m is illustrated in Fig. 4.3, which shows the emulator predictions for each statistic as Ω_m is varied individually while all other parameters are held fixed at their mean values. The colour indicates the values of the cosmological parameters, which range from the minimum to the maximum of the hypercube, as also indicated by the colour bar. In Appendix 4.D, we show similar figures for h, w_0, w_a, and Ω_b. Fig. 4.3 illustrates different dependencies on Ω_m. The emulator predictions across different Ω_m values result in changes of nearly two orders of magnitude in the amplitudes of the redshift and height distributions, and a factor of a few in the angular clustering amplitude, with the curves spanning the full dynamic range of the axes in all three panels. Out of the ten cosmological parameters, within the hypercube range, Ω_m shows the largest deviations across each of the three statistics. The amplitude of the redshift distribution is set by the number of peaks above the SNR > 3 threshold that is applied to the height distributions. Additionally, the redshift distribution shows a secondary dependence, as Ω_m also changes its shape. Although the distribution always peaks at the same redshift, which is close to the redshift at which the lensing kernel peaks given the chosen source redshift distribution, smaller values of Ω_m preferentially lead to a smaller fraction of lower redshift peaks, which could be due to the fact that as Ω_m decreases, Ω_Λ increases to maintain a spatially flat universe. Therefore, late-time density perturbations are increasingly suppressed, as in the limit of DE domination, the linear growth factor asymptotically approaches a constant, halting further structure growth.

For low values of Ω_m, the angular clustering signal increases monotonically with Ω_m. However, as Ω_m is increased beyond 0.3, the clustering signal shows smaller relative changes than the other two statistics and the signal at $\theta > 0.3$ deg even starts to decrease, while the smaller scales still show the same monotonic behaviour. This non-monotonic behaviour could explain why the clustering signal is not able to provide precise constraints on Ω_m with the 1D posterior even being multimodal. However, because the clustering constraint reaches the imposed prior boundary on Ω_m, these constraints should be interpreted with caution.

The amplitude of the primordial power spectrum, $\ln(10^{10}A_s)$, sets the overall normalisation of the matter power spectrum, directly controlling the abundance and masses of the structures that generate the high-SNR WL peaks. A larger value of A_s increases the amplitude of density fluctuations, leading to more massive haloes and a higher number density of high-SNR peaks. The constraint on $\ln(10^{10}A_s)$ is driven by the angular clustering and peak height distribution, as a higher $\ln(10^{10}A_s)$ produces more WL peaks at fixed κ and a more strongly clustered large-scale structure, leading to a stronger two-point signal and more high-SNR peaks.

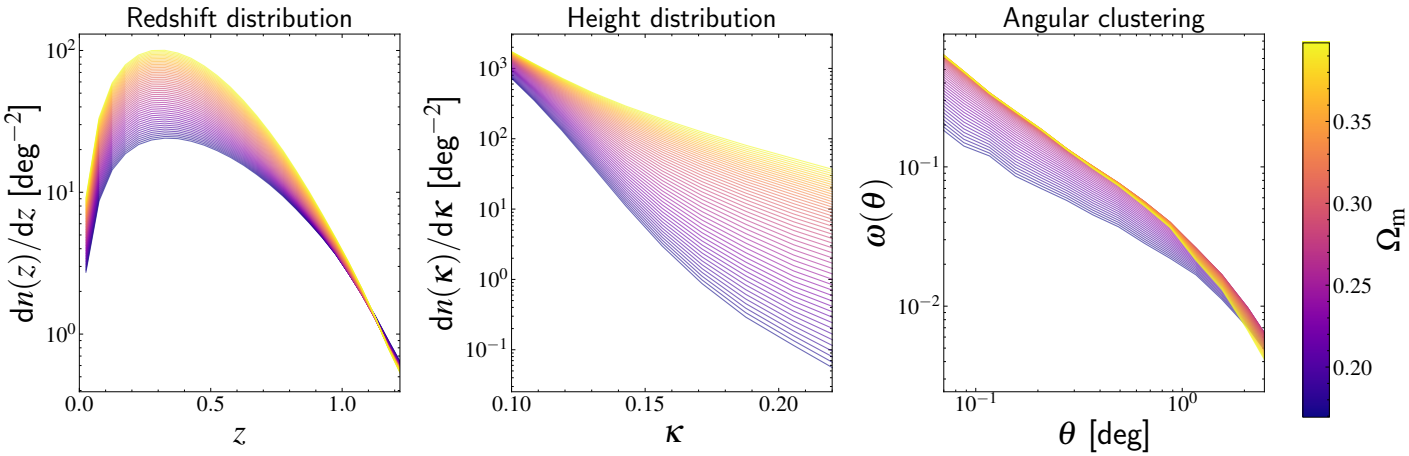


Figure 4.3: Emulator predictions for the redshift distribution (left), height distribution (middle), and angular clustering (right) of SNR > 3 WL peaks at 1 arcmin smoothing for varying Ω_m, as indicated by the colour bar, while keeping all other parameters fixed to their mean values within the hypercube. All three statistics are impacted. Larger values of Ω_m increase the overall number of peaks at fixed WL convergence and give rise to peaks preferentially corresponding to redshifts where the WL kernel peaks. Below 1 deg the angular clustering signal shows a monotonic trend of larger Ω_m resulting in a stronger clustering signal, but at larger angular scales this trend reverses.

Dark energy equation of state

In our forecast, the CPL equation-of-state parameters of DE, w_0 and w_a, affect the inference through their effects on the expansion history, which enter the lensing kernel, and through their impact on the WL convergence through changes in geometric and growth rates. The DE sets the late-time expansion history, where a less negative w_0 corresponds to stronger dark energy dominance, which both suppresses the growth of structure and alters the angular diameter distances to the lens and source planes. Because the simulations do not model dark energy perturbations, the effect of evolving DE parameters on the growth rate is purely through the expansion rate. Together, these effects modify the WL convergence value of a WL peak generated by a halo of fixed mass. In particular, the amplitude of the integrated convergence (Eq. 4.7) along the line of sight scales with the product of the matter density contrast, the comoving distances, and the kernel $W(\chi_i)$, all of which depend on w_0 and w_a.

As, among the three WL peak statistics, the redshift distribution is the only statistic that directly probes a signal over cosmic time, it is unsurprising that it dominates the constraining power on w_0 and w_a in our forecast. The forecast yields $w_0 = -0.98 \pm 0.06$ and $w_a = -0.15 \pm 0.24$. This represents a 6 per cent relative precision on w_0 while w_a remains poorly constrained with $\sigma_{w_a} = 0.24$. The parameter w_a controls the time-evolution of the equation of state and is generally less well constrained than w_0. This is because the constraining signal for w_a requires resolving the differential evolution of the dark energy contribution across redshift, which remains more uncertain in our forecast.

There is a strong degeneracy direction across all $w_0 - w_a$ constraints, as the two parameters affect the forecast in a very similar way. This is also reflected by the figures in Appendix 4.D showing the impact of varying the two parameters individually on the WL peak statistics.

Hubble parameter

The Hubble parameter, h, enters the WL peak statistics most directly through the physical matter density, which scales as $\Omega_\mathrm{m}h^2$, and sets the overall amplitude of the WL convergence (Eq. 4.9). As a consequence, h and Ω_m show a negative correlation that is visible in the 2D contour for the height distribution and angular clustering in Fig 4.9, as a larger value of h can compensate for a lower value of Ω_m to generate a peak of the same WL convergence value. This anti-correlation is primarily present in the posteriors corresponding to the height distribution and angular clustering, which therefore struggle to constrain h independently of Ω_m. The redshift distribution is able to break this degeneracy as it provides a tight estimate on Ω_m, such that the combination is also able to probe h, possibly through the impact of Ω_m on the geometry that is being probed by the redshift distribution. The qualitative impact of h on the angular clustering and height distribution

is very similar to that of Ω_m, as shown in Fig. 4.10, albeit with smaller absolute changes. Regarding the redshift distribution, h primarily impacts its amplitude, whereas Ω_m also affects the shape, illustrating how including the redshift distribution helps break the $\Omega_m - h$ degeneracy.

Baryon density

In the hypercube, Ω_m is varied individually and Ω_b is varied through changes in the ratio Ω_b/Ω_m. Ω_b at fixed Ω_b/Ω_m then impacts the forecast through its impact on the expansion history. At fixed Ω_m, the expansion history is not affected, and changes in Ω_b/Ω_m, instead, likely affect the inference through changes in the primordial power spectrum. Individually, none of the three statistics provides a tight constraint on Ω_b. However, when combining all three statistics, the constraint on Ω_b improves drastically. The physical origin of this improvement is not straightforward to identify, but it reflects the breaking of a degeneracy in the full 10D parameter space that no single statistic can resolve on its own. We note that Ω_b is, in any case, already well constrained independently through Big Bang nucleosynthesis (e.g. Fields et al., 2020), and a strong external prior on this parameter could help break the residual degeneracies even further, potentially tightening the constraints on the remaining cosmological parameters in the process.

Neutrino mass, scalar spectral index, and beyond w_0w_aCDM parameters

Several parameters in the hypercube are only weakly constrained by our fiducial forecast, and the marginalised posteriors recover much of the prior range. The sum of neutrino masses $\sum m_\nu c^2$ is not well constrained in our combined forecast compared to competitive results from measurements of, for example, the CMB combined with DESI BAO, which report constraints on the neutrino mass with a marginalised error of 0.078 eV for a w_0w_aCDM+Σm_ν inference (Elbers et al., 2025a). The constraining power, however, is still competitive with CMB-independent constraints from DESI. Using the same simulations as us, Wang et al. (2026) showed that persistence strips, a formalism that segments Betti curves by topological persistence, can reach a neutrino mass constraint with an uncertainty down to 0.05 eV.

The scalar spectral index n_s and its running α_s affect the shape of the primordial power spectrum and, hence, the matter power spectrum on small scales, which ultimately influence the mass function of haloes and, in turn, the high-SNR peaks. However, given that the range over which these parameters are varied in the hypercube is not very large relative to existing CMB constraints, and the constraining power of the WL peak statistics is limited, we fail to constrain these parameters within our fiducial forecast.

Similarly, the DDM decay rate $\Gamma_{\rm dcdm}$ is only weakly constrained, and the marginalised posterior recovers the half-Gaussian prior we impose. This is somewhat surprising, as in Broxterman et al. (2025) we found that the amplitude of the FLAMINGO variations with DDM rates of 0.12 and 0.24 H_0/h reduced the amplitude of the redshift distribution by factors of $\approx$ 2 and $\approx$ 4, respectively. Although the latter is outside the range of the hypercube used in this work, naively, given the change in the number of peaks above the SNR = 3 threshold in the 0.12 H_0/h run, we would still expect an additional constraint beyond the prior in this work. The two analyses differ in their peak-height cut and noise properties, although the exact reason we do not recover stronger constraints on $\Gamma_{\rm dcdm}$ is unclear, it might be the fact that we already impose a strong prior on the value of $\Gamma_{\rm dcdm}$ and the small variations within this prior do not impact the analysis much. The prior on $\Gamma_{\rm dcdm}$ is more than 10 times stronger than the weakest DCDM scenario considered in the FLAMINGO suite. For $\Gamma_{\rm dcdm}$ = 1 km/s/Mpc, which is the width of the half-Gaussian prior we apply, the effect on the linear power spectrum is less than 1 per cent and the effect on the expansion rate is less than 0.1 per cent. An alternative explanation is that our emulators emulate $\Omega_{\rm m}$, i.e. the real present-day value of $\Omega_{\rm b} + \Omega_{\rm cdm} + \Omega_\nu$ and not $\Omega_{\rm m,noDDM}$, which is the parameter that is varied in the hypercube. Some of the sensitivity to $\Gamma_{\rm dcdm}$ may therefore be absorbed by $\Omega_{\rm m}$.

4.5.3 Complementarity of WL peak statistics

One of the key results of this work is the complementarity of the three different WL peak statistics. Fig. 4.2, and in more detail, the 2D corner plot in Appendix 4.C, demonstrates that the combination of the WL peak height distribution, redshift distribution, and angular clustering significantly outperforms each individual statistic for the majority of parameters. Particularly the constraints on $\Omega_{\rm b}$, h, w_a, $\sum m_\nu c^2$, and $\ln(10^{10} A_{\rm s})$ improve significantly compared to any individual statistic.

The height distribution is directly sensitive to the number density of massive haloes as a function of their mass. Individually, the height distribution does not impose strong constraints on any cosmological parameters. Since the height distribution measures the WL convergence values by integrating over the entire source redshift distribution, it averages over the complete growth history, which limits its ability to constrain parameters that affect the growth history in different ways but have similar effects on the amplitude of the WL signal.

Adding the redshift distribution allows us to break some of these degeneracies by probing the evolution of structure formation. Individually, the redshift distribution is only able to constrain well $\Omega_{\rm m} - w_0$. However, the corner plot shows that combining the statistics helps break degeneracies. Individually, none of the statistics can probe the neutrino mass, but together they can place a clear

upper bound, albeit still not competitive with other cosmological probes. Angular clustering provides a qualitatively different source of information by probing the spatial distribution of peaks on the sky. Consequently, it provides clear constraints on $\ln(10^{10}A_s)$.

Davies et al. (2022) found that the clustering of WL peaks is a factor ≈ 3 more sensitive to h and w_0 than the peak height distribution is. For w_0, we see a qualitatively similar improvement between the red and blue 1D posteriors. However, the redshift distribution outcompetes the height distribution and angular clustering for these parameters. For h, we do not see the same improvement. If anything, the height distribution seems to provide slightly better results. The comparison is complicated by the fact that, in our 10D inference, we cannot constrain these parameters within the hypercube bounds, so the prior bounds are therefore important. Also, Davies et al. (2022) combine the clustering of peaks selected at different heights, which they show significantly improves the constraints compared to clustering of peaks at a single cut-off value.

4.5.4 Comparison to shear two-point correlation functions

As our analysis is idealised, since we do not include marginalisation over nuisance parameters, it is difficult to quantitatively compare our constraints to those of other works, let alone to the precision of current inferences or requirements for *Euclid*. Therefore, to more straightforwardly assess the benefits and downsides of our WL peak forecast, we compare our results to those of a similarly idealised forecast using WL shear two-point correlation functions (2PCFs), $\xi_\pm$, which are the standard statistic used by current cosmic shear surveys. The 2PCFs are computed from the same simulations and WL convergence maps and use the same smoothing, priors, and covariance matrix realisations as the fiducial setup. Fig. 4.4 shows the marginalised 2D posterior distributions for all 10 cosmological parameters for the $\xi_\pm$ 2PCFs (blue), the combined WL peak statistics (black), and the combination of all four statistics (orange), i.e. the shear 2pt functions plus the three peak statistics. In general, the combination of WL peak statistics outperforms the $\xi_\pm$ 2PCFs. In reality, the cosmic shear signal is measured across different tomographic bins, meaning our forecast on the constraining power of the 2PCFs is a conservative estimate. Additionally, the cosmic shear measurements are usually combined with galaxy clustering and galaxy-galaxy lensing in a 3×2pt analysis, yielding better constraints than the cosmic shear two-point signal alone. A full comparison to the 3×2pt, or even 6×2pt, forecast, however, is beyond the scope of this work.

For the parameters Ω_m, Ω_b, w_0, and w_a, the combination of WL peak statistics provides a tighter constraint than the $\xi_{+/-}$, while the constraints on h, $\ln(10^{10}A_s)$, and Γ_{dcdm} are of a similar level. Only for the parameters $\Sigma m_\nu c^2$, n_s, and α_s, which were poorly constrained in the WL peak forecast, do the $\xi_{+/-}$ outperform the WL

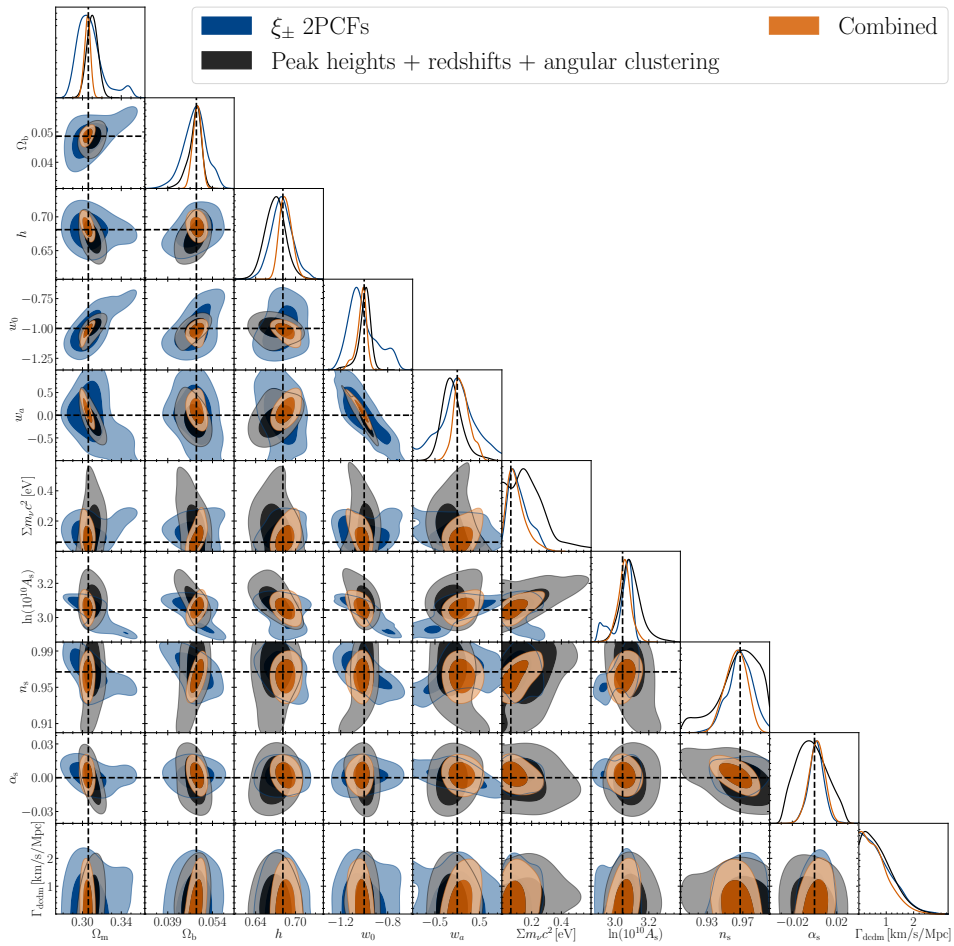


Figure 4.4: Corner plot of the cosmological parameter forecast using the WL two-point correlation functions (blue) or angular clustering, redshifts and heights of WL peaks (black) of SNR > 3 peaks or a combined inference using all 4 statistics (orange) at 1 arcmin smoothing. The contours indicate the 1- and 2σ credible intervals. The combination of WL peak statistics outperforms the $\xi_{+/-}$ 2PCFs for $\Omega_\mathrm{m}, \Omega_\mathrm{b}, w_0$, and w_a, while the constraints on h, $\ln(10^{10}A_\mathrm{s})$ and Γ_dcdm are similar and the remaining parameters, $\Sigma m_\nu c^2, n_\mathrm{s}$, and α_s are probed best by the $\xi_{+/-}$ 2PCFs.

peaks. The constraints on these parameters, however, remain worse than those obtained from other cosmological observables, such as the CMB (see e.g. Planck Collaboration et al., 2020). Whereas the high-valued WL peaks are sensitive to the most massive haloes and structures, the 2PCFs are instead sensitive to the shape of the matter power spectrum across a wide range of scales, which is why the $\xi_{+/-}$ constrains these parameters better (see, e.g. Hamana et al., 2004; Heitmann et al., 2009; Yang et al., 2025; Broxterman et al., 2025).

The joint forecast tightens the contours even further, particularly for Ω_m, where some parameter degeneracy seems to be broken as the WL peaks and $\xi_{+/-}$ show different degeneracies in, for example, the $\Omega_m - h$ and $\Omega_m - \ln(10^{10} A_s)$ subpanels.

The results are qualitatively consistent with Davies et al. (2022), who also find that WL peaks can probe S_8 and w_0 more precisely than the shear-2PCFs. However, they find a similar level of constraining power on Ω_m and that the shear-2PCFs are able to probe h better than the peak statistics. This could be explained by the fact that they do not include the redshift distribution of WL peaks, which drives our constraints on Ω_m. As expected, they generally find that combining the statistics leads to even tighter constraints.

Overall, these results suggest that WL peaks and 2PCFs are complementary probes, as the peaks carry non-Gaussian information that helps break degeneracies in the 2PCF posterior, while the 2PCFs still carry constraints on parameters to which the peak statistics are less sensitive. Their combination, therefore, yields a more powerful cosmological inference than either statistic alone, motivating a joint analysis for future *Euclid*-like surveys, which can be further improved by including additional non-Gaussian statistics (see e.g. Euclid Collaboration: Ajani et al., 2023).

4.5.5 Impact of the smoothing scale

In this section, we explore the impact of the smoothing scale on the cosmological constraints. We chose our fiducial smoothing scale based on the findings of Liu et al. (2015), who found it to be the most constraining for their inference of the WL peak-height distribution. However, they already showed that combining scales may improve the results, leaving a more dedicated analysis to future work.

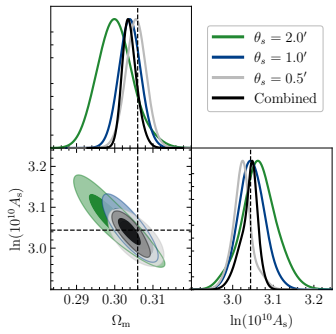
Physically, a different smoothing scale corresponds to a different physical size at each lens redshift. In order to measure an object with the highest significance, the physical aperture corresponding to the smoothing scale should be similar to the object's size. The smoothing will thus affect which WL peaks are detected and at what significance level. By combining scales that are better suited for extracting smaller high-redshift or larger lower-redshift objects, more information could possibly be probed. Additionally, for the smallest scales, extra information may

be obtained from different radial profile parts of haloes, as a single low-redshift object may generate multiple peaks.

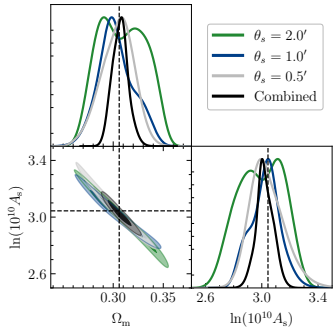
We fix eight cosmological parameters to their true values and infer only the remaining two. Consequently, the absolute constraining power may differ from the previous figures. We focus on the $\Omega_\mathrm{m} - \ln(10^{10} A_\mathrm{s})$ and $w_0 - w_a$ planes because these are the parameter combinations that upcoming WL inferences will likely be most interested in. At the same time, because we did not re-optimise our emulators for different smoothing scales or for this 2D parameter forecast, we found the results to be less stable to small variations in the emulation set-up. We therefore decided to add the emulator uncertainty as an additional noise component to the covariance matrix $\mathbf{C}$ (Eq. 4.15). Following the same leave-one-out approach described in Appendix 4.A, we estimate an emulator error covariance matrix, Σ_emu, from the residuals between the emulator predictions and the true measurements of the withheld hypercube nodes. The total covariance used in this section is then computed following the combined likelihood function for a multivariate t and Gaussian distribution from Heavens et al. (2026). The additional term increases the uncertainty, preventing the 2D forecast from becoming artificially overconfident due to emulator imperfections.

From top to bottom, Figs. 4.5 and 4.6 show the marginalised 2D posterior distributions for the redshift distribution, height distribution, and angular clustering, respectively, for smoothing scales of $\theta_s = 0.5$ (grey), 1.0 (blue), and 2.0 (green) arcmin, as well as the forecast using the three scales combined (black). The results are further quantified in Table 4.2. There, for each statistic at the different smoothing scales, as indicated by the first column, the FoM in the $\Omega_\mathrm{m} - \ln(10^{10} A_\mathrm{s})$ and $w_0 - w_a$ planes are listed in the second and fourth columns, respectively. We only interpret the relative changes in the FoMs due to our idealised forecasts. Similarly, the corresponding FoB values are listed in the third and fifth columns.

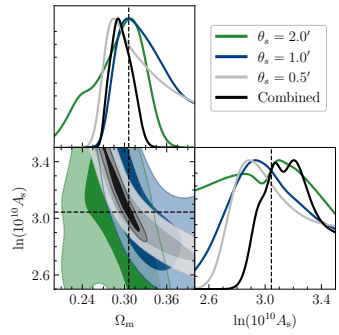
The FoB at the 68th-percentile confidence level for this 2D inference is 1.52. Forecasts with a bias below this threshold are therefore less biased than the commonly used 1σ value to quantify the bias for a single parameter. The FoB values are below the 1.52 threshold in almost every case. Only for the $w_0 - w_a$ forecast using the redshift distribution at $\theta_\mathrm{s} = 2.0$'the FoB is 2.35.



(a) Redshift distribution

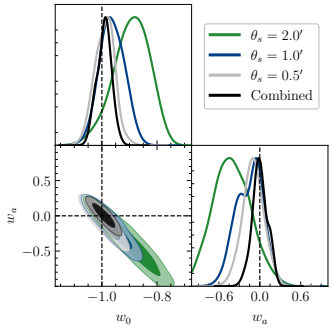


(b) Height distribution

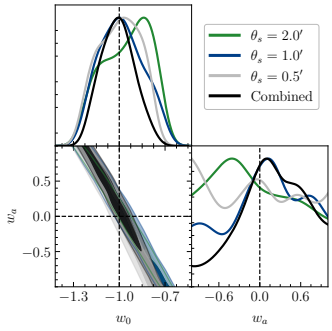


(c) Angular clustering

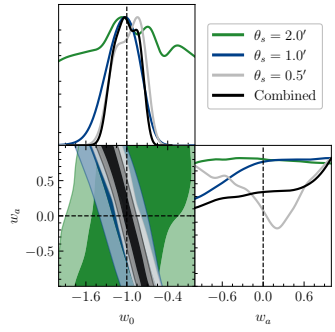
Figure 4.5: Posteriors for the 2 dimensional forecast for $\Omega_m - \ln(10^{10}A_s)$ using (a) the redshift distribution (top), (b) the height distribution (middle) or (c) the angular clustering (bottom) for 0.5′(grey), 1.0′(blue), 2.0′(green), or the combination of all three (black) smoothing scales. The redshift distribution shows the clearest improvement when combining multiple smoothing scales. The height distribution shows a weaker improvement, whereas the angular clustering shows the characteristic WL banana shape.



(a) Redshift distribution



(b) Height distribution



(c) Angular clustering

Figure 4.6: Same as Fig. 4.5 but for $w_0 - w_a$ posteriors. For the redshift distribution, a smaller smoothing scale yields tighter constraints, while combining smoothing scales yields significantly tighter constraints. The height distribution shows a weaker improvement, and even in the combined case the angular clustering remains insensitive to the time evolution of DE.

Table 4.2: Figure of merit (FoM) and figure of bias (FoB) for the $\Omega_m - \ln(10^{10}A_s)$ and $w_0 - w_a$ parameter planes when all other cosmological parameters are fixed while varying the smoothing scale for a forecast using the redshift distribution, the height distribution, or angular clustering, at smoothing scales of $\theta_s = 0.5'$, $1.0'$, or $2.0'$, and for the combination of all three scales. The FoM and FoB corresponding to contours that hit the prior bounds, as shown in Figs. 4.5 & 4.6, should be treated carefully, as the prior truncation may bias the inferred covariance used to compute the FoM and FoB. The redshift distribution provides the tightest constraint on both parameter combinations. The FoM of all statistics and parameter combinations typically increases by a factor ≈ 2 when combining the different scales.

	$\Omega_m - \ln(10^{10}A_s)$		$w_0 - w_a$	
Statistic and smoothing scale	Figure of merit	Figure of bias	Figure of merit	Figure of bias
Redshift distribution at $\theta_s = 0.5'$	18780.5	0.87	347.2	0.90
Redshift distribution at $\theta_s = 1'$	17850.2	0.75	243.8	0.73
Redshift distribution at $\theta_s = 2'$	10744.9	1.31	151.9	2.35
Redshift distribution at $\theta_s = 0.5' + \theta_s = 1' + \theta_s = 2'$	38363.5	1.14	649.0	0.48
Height distribution at $\theta_s = 0.5'$	2868.8	0.26	59.6	1.24
Height distribution at $\theta_s = 1'$	2842.6	0.28	57.5	0.21
Height distribution at $\theta_s = 2'$	2359.0	0.07	39.8	1.05
Height distribution at $\theta_s = 0.5' + \theta_s = 1' + \theta_s = 2'$	6623.2	0.75	80.9	0.28
Angular clustering at $\theta_s = 0.5'$	251.2	0.50	16.4	0.14
Angular clustering at $\theta_s = 1'$	153.5	0.43	10.2	0.31
Angular clustering at $\theta_s = 2'$	105.2	0.26	3.2	0.41
Angular clustering at $\theta_s = 0.5' + \theta_s = 1' + \theta_s = 2'$	1344.7	0.95	25.1	0.11

The three statistics show similar improvements when combining the three smoothing scales. For the redshift distribution, combining all three scales improves the FoM by a factor of 2.15 and 2.66 for $\Omega_\mathrm{m} - \ln(10^{10} A_\mathrm{s})$ and $w_0 - w_a$, respectively, compared to only using our fiducial 1 arcmin scale. The absolute improvement is the largest of the three statistics. The height distribution shows a relatively similar improvement for $\Omega_\mathrm{m} - \ln(10^{10} A_\mathrm{s})$ of 2.33, while for $w_0 - w_a$ the combined FoM improves by a factor of 1.41 relative to the fiducial scale. The relative gain is similar to that of the redshift distribution, even though the absolute constraining power of the height distribution on $w_0 - w_a$ remains weaker than that of the redshift distribution at every individual scale. The angular clustering shows the largest relative improvement of all three statistics, with the combined FoM exceeding the fiducial-scale value by a factor of 8.76 for $\Omega_\mathrm{m} - \ln(10^{10} A_\mathrm{s})$ and 2.46 for $w_0 - w_a$. As mentioned before, Davies et al. (2022) found that combining peaks selected at multiple SNR thresholds substantially improved angular clustering constraints. Instead, we compared peaks selected from the same SNR threshold but from different smoothing scales, which qualitatively probe the same information. We also find that the results improve significantly when combining multiple, differently selected angular clustering distributions. However, in terms of absolute constraining power, the angular clustering remains the poorest-performing statistic in both parameter planes.

Comparing the different statistics at the fiducial 1 arcmin smoothing scale, the redshift distribution dominates the constraining power in both parameter planes, exceeding the height distribution by a factor of 6 and the angular clustering by a factor of 116 in FoM for $\Omega_\mathrm{m} - \ln(10^{10} A_\mathrm{s})$, and by factors of 4 and 24, respectively, for $w_0 - w_a$.

Regarding the sensitivity to the smoothing scale itself, each of the three statistics shows a monotonic trend in which the smallest scale yields the tightest constraints in both parameter planes. Given the number of covariance matrix realisations, it is not possible to include an extra scale in the forecast as the number of data vector bins for the redshift distribution would exceed the number of covariance matrix realisations, even though this might further improve the constraints. However, forecasts of each of the three statistics in both parameter planes, with 5 arcmin smoothing, show that constraints weaken further (not shown), consistent with the trend that smaller smoothing scales yield better constraints.

Compared to the findings of Liu et al. (2015), who identified ≈1 arcmin as close to the optimal smoothing scale for a CFHTLenS-depth inference based solely on the peak number density, our results suggest that, for a *Euclid* inference using SNR>3 peaks, a slightly smaller scale is preferred, likely due to (a combination of) the better angular resolution, lower noise levels and higher source number density. The smallest smoothing scale we consider corresponds to only 1.16 pixels for the $N_\mathrm{side} = 8192$ maps that are used. As Healpix' functionality, which we need to smooth and rotate our mass maps, is limited by 32-bit

indexing, we did not explore smaller scales, even though they may provide even stronger constraints.

The statistics exhibit qualitatively different directions of degeneracy. The redshift and height distributions both show negatively correlated degeneracies between Ω_m and $\ln(10^{10}A_\mathrm{s})$ at every smoothing scale, whereas for the angular clustering forecast, the $\theta_\mathrm{s} = 0.5$ and 1 arcmin $\Omega_\mathrm{m} - \ln(10^{10}A_\mathrm{s})$ posteriors show a banana-shaped degeneracy that is not apparent in the fiducial 10D forecast, which is similar to the banana-shaped posteriors typically found in WL two-point cosmic shear inferences in the $\sigma_8 - \Omega_\mathrm{m}$ plane. In our inference, the curvature is a consequence of the non-monotonic Ω_m dependence of the clustering signal, as shown before in Fig. 4.3. The clustering amplitude increases with Ω_m only up to $\Omega_\mathrm{m} \approx 0.3$, beyond which the impact decreases and a reversed trend is even observed at large scales. At the same time, within our hypercube range, the dependence of clustering amplitude on $\ln(10^{10}A_\mathrm{s})$ is always monotonic, with larger values of $\ln(10^{10}A_\mathrm{s})$ resulting in a stronger clustering signal (not shown). Despite being noisier, the 2′scale shows a different degeneracy than the other two smoothing scales. Consequently, when combining the scales, the slightly different degeneracies cause a factor ≈ 9 increase in the FoM compared to the fiducial smoothing scale. It is unclear whether this represents a different physical behaviour or is instead a result of some part of the pipeline. Similar, somewhat confusing behaviour can be seen in the $w_0 - w_a$ posteriors for the angular clustering forecast at 2′, which appear to reverse the degeneracy direction, albeit remaining practically unconstrained within the prior ranges.

More generally, our exploration of sensitivity to the smoothing scale is illustrative rather than exhaustive. Although we typically find that the smaller smoothing scale yields the tightest constraint, we have considered only three discrete scales and only in two restricted parameter planes, and we have ignored baryonic physics, which potentially impacts the smallest scales the strongest. A more complete analysis that combines multiple smoothing scales and multiple statistics across the full 10D forecast is left for future work. Similarly, we did not explore the possibility of mixing different scales for the different statistics.

4.6 Conclusions

We have investigated the potential of using WL peak properties to constrain cosmology in a *Euclid*-like setup that indirectly aims to exploit the cosmology dependence of the mass, angular clustering, and redshift distribution of massive structures. We have done so by forward-modelling the final expected signal in a new DMO hypercube, a suite of simulations spanning a ten-dimensional cosmological parameter space varying the matter density, evolving dark energy, neutrino mass, and beyond $w_0 w_a$CDM parameters that control the running of the scalar spectral index and the decay rate of cold dark matter. We have run an additional suite of

80 ΛCDM DMO simulations within the same simulation framework to estimate the covariance matrix at the fixed D3A cosmology.

Within these simulations, we generate WL convergence maps from which we extract the local maxima, i.e. WL peaks. We study the cosmological constraining potential of the different WL peak properties, namely (the combination of) their angular clustering, redshift distribution, and height distributions. We find that different WL peak statistics probe different regions of the cosmological space. For our fiducial set-up that uses SNR > 3 peaks measured from a map smoothed with 1 arcmin, the WL peak redshift distribution provides the tightest constraints on the matter density, Ω_m, and on the dark energy equation-of-state parameters, w_0 and w_a. The angular clustering and peak-height distribution jointly constrain best the amplitude of the primordial power spectrum, $\ln(10^{10}A_\mathrm{s})$. The combination of different statistics can also probe the Hubble parameter, h, and the baryon matter density, Ω_b, through the breaking of parameter degeneracies. None of the considered WL peak statistics, even when combined, provide strong constraints on the dark matter decay rate, the neutrino mass, or the amplitude and running of the scalar spectral index.

We interpret the different cosmological constraining power by considering three pathways through which cosmological forecasts are affected by changes in cosmology, namely through direct impact on the matter density, changes in geometry, or changes in structure formation. Although it is difficult to isolate the effect of a single channel, we find that all channels likely contribute to the constraints, given that the different statistics exhibit distinct degeneracies and sensitivities to cosmological parameters.

We additionally compare the constraining power to that of the commonly used two-point cosmic-shear correlation functions. We find that for most parameters, the combination of WL peak statistics provides equal or better constraining power than the $\xi_{+/-}$. Only for the neutrino mass and the (running of the) scalar spectral index do we find that the shear-2PCFs outperform the WL peaks, which do not constrain these parameters well. When combining both, the constraints improve further.

We additionally explore the impact of the smoothing scale to explore the optimal scale and complementarity of different scales in the 2D $\Omega_\mathrm{m} - \ln(10^{10}A_\mathrm{s})$ and $w_0 - w_a$ planes by comparing cosmological forecasts for peaks selected from maps smoothed with 0.5, 1, and 2 arcmin. For both parameter planes, we find the redshift distribution to provide the tightest constraints, with its FoM being a factor $\approx 10-100$ larger than the corresponding FoMs for the angular clustering or height distribution. When combining different smoothing scales, which select partially independent sets of objects with different masses and redshifts, we find the FoMs to typically improve by a factor ≈ 2.

The cosmological forecast setup of this paper is idealised, as it ignores the impact of baryonic physics, intrinsic alignment, source clustering, reduced shear,

and nuisance parameters that are expected to significantly impact the results. Although the impact of some of these effects and systematics is expected to be minor or qualitatively different for the different WL peak statistics we considered, future analyses will need to better account for these effects to obtain precise and accurate cosmological constraints from upcoming data from *Euclid*, *Roman*, and *Rubin*.

In addition, we restricted our analysis to WL peaks and did not consider other higher-order WL statistics in detail, although including them can further improve the constraints (see e.g. Marques et al., 2024, for the inclusion of WL minima).

4

Acknowledgements

This work was supported by the Science and Technology Facilities Council (grant number ST/Y002733/1). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 769130). 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. We have used ASTROPY (Astropy Collaboration et al., 2022), HEALPY (Zonca et al., 2019), MATPLOTLIB (Hunter, 2007), NUMPY (Harris et al., 2020), and SWIFTSIMIO (Borrow & Borrisov, 2020) in our analysis.

Data Availability

The data supporting the figures in this article are available upon reasonable request to the corresponding author. The hypercube will be presented in McCarthy et al. (in preparation) and will eventually be made public.

References

Abbott T. M. C., et al., 2022, Phys. Rev. D, 105, 023520
Abdul Karim M., et al., 2025, Phys. Rev. D, 112, 083515
Adame A. G., et al., 2025, J. Cosmology Astropart. Phys., 2025, 028
Allys E., Marchand T., Cardoso J.-F., Villaescusa-Navarro F., Ho S., Mallat S., 2020, Phys. Rev. D, 102, 103506
Astropy Collaboration et al., 2022, ApJ, 935, 167
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bilicki M., et al., 2021, A&A, 653, A82
Borrow J., Borrisov A., 2020, The Journal of Open Source Software, 5, 2430
Bradač M., Schneider P., Lombardi M., Erben T., 2005, A&A, 437, 39
Broxterman J. C., et al., 2024, MNRAS, 529, 2309
Broxterman J. C., Schaller M., Hoekstra H., Schaye J., McGibbon R. J., Forouhar Moreno V. J., Kugel R., Elbers W., 2025, MNRAS, 538, 755
Chen K.-F., et al., 2025, The Open Journal of Astrophysics, 8, 2
Cheng S., Ting Y.-S., Ménard B., Bruna J., 2020, MNRAS, 499, 5902
Chevallier M., Polarski D., 2001, International Journal of Modern Physics D, 10, 213
Chiu I.-N., et al., 2024, The Open Journal of Astrophysics, 7, 90
DES Collaboration et al., 2026a, arXiv e-prints, p. arXiv:2601.14559
DES Collaboration et al., 2026b, arXiv e-prints, p. arXiv:2605.27221
Davies C. T., Cautun M., Li B., 2019, MNRAS, 488, 5833
Davies C. T., Cautun M., Giblin B., Li B., Harnois-Déraps J., Cai Y.-C., 2022, MNRAS, 513, 4729
de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, Experimental Astronomy, 35, 25
Dietrich J. P., Hartlap J., 2010, MNRAS, 402, 1049
Elbers W., et al., 2025a, Phys. Rev. D, 112, 083513
Elbers W., et al., 2025b, MNRAS, 537, 2160
Euclid Collaboration: Ajani V., Baldi M., Barthelemy A., et al., 2023, A&A, 675, A120
Euclid Collaboration: Blanchard A., Camera S., Carbone C., et al., 2020, A&A, 642, A191
Euclid Collaboration: Cañas-Herrera G., Goh L. W. K., Blot L., et al., 2025, A&A, in press, https://doi.org/10.1051/0004-6361/202556861, p. arXiv:2510.09153
Euclid Collaboration: Mellier Y., Abdurro'uf Acevedo Barroso J., et al., 2025, A&A, 697, A1
Euclid Collaboration: Vinciguerra S., Bouchè F., Martinet N., et al., 2026, A&A, 707, A235
Feldbrugge J., van Engelen M., van de Weygaert R., Pranav P., Vegter G., 2019, J. Cosmology Astropart. Phys., 2019, 052
Fields B. D., Olive K. A., Yeh T.-H., Young C., 2020, J. Cosmology Astropart. Phys., 2020, 010
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Gatti M., et al., 2024, MNRAS, 527, L115
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Grandón D., Marques G. A., Thiele L., Cheng S., Shirasaki M., Liu J., 2024, Phys. Rev. D, 110, 103539
Hahn O., Michaux M., Rampf C., Uhlemann C., Angulo R. E., 2020, MUSIC2-monofonIC: 3LPT initial condition generator, Astrophysics Source Code Library, record ascl:2008.024 (ascl:2008.024)

Hamana T., Takada M., Yoshida N., 2004, MNRAS, 350, 893
Harris C. R., et al., 2020, Nature, 585, 357
Heavens A., Whiteway L., Sellentin E., 2026, The Open Journal of Astrophysics, 9, 65025
Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, ApJ, 705, 156
Heydenreich S., Brück B., Harnois-Déraps J., 2021, A&A, 648, A74
Hildebrandt H., et al., 2020, A&A, 633, A69
Hoekstra H., Jain B., 2008, Annual Review of Nuclear and Particle Science, 58, 99
Hunter J. D., 2007, Computing in Science and Engineering, 9, 90
Jain B., Van Waerbeke L., 2000, ApJ, 530, L1
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Kodwani D., Alonso D., Ferreira P., 2019, The Open Journal of Astrophysics, 2, 3
Kratochvil J. M., Haiman Z., May M., 2010, Phys. Rev. D, 81, 043519
Kugel R., et al., 2023, MNRAS, 526, 6103
Kugel R., Schaye J., Schaller M., McCarthy I. G., Braspenning J., Helly J. C., Forouhar Moreno V. J., McGibbon R. J., 2024, MNRAS, 534, 2378
Kugel R., Schaye J., Schaller M., Forouhar Moreno V. J., McGibbon R. J., 2025, MNRAS, 537, 2179
LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201
Lahav O., Liddle A. R., 2024, arXiv e-prints, p. arXiv:2403.15526
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Lee M. E., Haiman Z., Pandey S., Genel S., 2026, ApJ, 996, 36
Lesgourgues J., 2011, arXiv e-prints, p. arXiv:1104.2932
Lewis A., 2025, J. Cosmology Astropart. Phys., 2025, 025
Li Z., Liu X., Fan Z., 2023, MNRAS, 520, 6382
Limber D. N., 1953, ApJ, 117, 134
Linder E. V., 2003, Phys. Rev. Lett., 90, 091301
Linke L., et al., 2026, arXiv e-prints, p. arXiv:2606.12389
Liu J., Haiman Z., Hui L., Kratochvil J. M., May M., 2014, Phys. Rev. D, 89, 023515
Liu J., Petri A., Haiman Z., Hui L., Kratochvil J. M., May M., 2015, Phys. Rev. D, 91, 063507
Lu T., Haiman Z., 2021, MNRAS, 506, 3406
Marian L., Smith R. E., Hilbert S., Schneider P., 2013, MNRAS, 432, 1338
Marini I., et al., 2024, A&A, 689, A7
Marques G. A., Liu J., Zorrilla Matilla J. M., Haiman Z., Bernui A., Novaes C. P., 2019, J. Cosmology Astropart. Phys., 2019, 019
Marques G. A., et al., 2024, MNRAS, 528, 4513
Michaux M., Hahn O., Rampf C., Angulo R. E., 2021, MNRAS, 500, 663
Miyazaki S., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460Z, doi:10.1117/12.926844
Pantos I., Perivolaropoulos L., 2026, Physics of the Dark Universe, 52, 102286
Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825
Planck Collaboration et al., 2020, A&A, 641, A6
Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning
Reischke R., et al., 2026, A&A, 709, A82
Rykoff E. S., et al., 2014, ApJ, 785, 104
Schaller M., et al., 2024, MNRAS, 530, 2378
Schaye J., et al., 2023, MNRAS, 526, 4978
Schneider P., van Waerbeke L., Mellier Y., 2002, A&A, 389, 729
Seitz C., Schneider P., 1995, A&A, 297, 287
Sellentin E., Heavens A. F., 2016, MNRAS, 456, L132
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
Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022

Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
The Dark Energy Survey Collaboration 2005, arXiv e-prints, pp astro–ph/0510346
Vedder C., Bakx T., Chisari N. E., Hoekstra H., Schaller M., 2026, MNRAS, 550, stag1008
Vinciguerra S., Martinet N., Gatti M., 2026, arXiv e-prints, p. arXiv:2607.17942
Wang S., Haiman Z., May M., 2009, ApJ, 691, 547
Wang J., Elbers W., Frenk C. S., Cole S., Yang X., McCarthy I. G., van de Weygaert R., 2026, MNRAS, 549, stag937
Weiss A. J., Schneider A., Sgier R., Kacprzak T., Amara A., Refregier A., 2019, J. Cosmology Astropart. Phys., 2019, 011
Williamson J., et al., 2026, arXiv e-prints, p. arXiv:2606.11309
Wright A. H., et al., 2025, A&A, 703, A158
Yang X., Kratochvil J. M., Wang S., Lim E. A., Haiman Z., May M., 2011, Phys. Rev. D, 84, 043529
Yang Y., Bird S., Ho M.-F., 2025, Phys. Rev. D, 111, 083529
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, The Journal of Open Source Software, 4, 1298
Zürcher D., Fluri J., Sgier R., Kacprzak T., Refregier A., 2021, J. Cosmology Astropart. Phys., 2021, 028

Appendix

4.A Emulator construction and validation

In this appendix, we provide additional details on the construction and accuracy of the emulators used for the forecasting. We use Gaussian process (GP) emulators that are built using SCIKIT-LEARN's GaussianProcessRegressor class (Pedregosa et al., 2011). All emulators use a Matérn-$\frac{3}{2}$ kernel with a trainable amplitude (see e.g. Rasmussen & Williams, 2006). We add a white-noise kernel to absorb stochasticity, as can be seen in, for example, the high-κ tail of the peak height distributions in Fig. 4.1.

The kernel hyperparameters are optimised by maximising the GP log-marginal likelihood, with the optimisation repeated 10 times with random restarts to avoid local maxima. The different WL peak summary statistics are emulated separately at each smoothing scale, and the resulting emulators are concatenated before forecasting. The different statistics and their corresponding emulator architectures are described below.

4.A.1 Redshift distribution

SNR > 3 peaks are selected and binned in redshift using 25 linearly spaced bins over $z \in [0, 1.25]$, with bin widths normalised to yield the density $\mathrm{d}N/\mathrm{d}z$. We chose this SNR cut because in Broxterman et al. (2025), we showed WL peaks above the corresponding κ-cut can be robustly assigned to a single massive object. Also, lower-valued peaks are more sensitive to the noise and instrument systematics, as discussed in Appendix B of Broxterman et al. (2024). In this way, we aim at a potentially robust cosmological inference. Note, however, that we apply stronger noise, so the SNR differs. The binning corresponds to the FLAMINGO light-cone redshift resolution and is similar to the expected redshift uncertainty on the massive objects that generate the high-valued WL peaks. For additional details, see the discussion in Section 3.5 of Broxterman et al. (2025). Before emulating the distributions, we apply a logarithmic transform $\ln(\mathrm{d}N/\mathrm{d}z + \epsilon)$ (with $\epsilon = 10^{-12}$) to stabilise the variance at low counts. The data vectors are then emulated using principal component analysis (PCA) separately for the shape of the redshift distribution and the normalisation, i.e., the total number of peaks above

the SNR > 3 threshold. We emulate the first 15 PCA components. When predicting the redshift distribution, the shape is reconstructed from an inverse PCA transform and multiplied by the amplitude prediction to yield the full data vector.

4.A.2 Height distribution

Peaks above the SNR threshold are binned into 10 logarithmically spaced SNR bins, with the upper bin edge set at the 99th percentile of the κ distribution across all training simulations. We do this because preliminary tests showed it yielded better predictions than including all high-end κ tail peaks, likely due to cosmic variance. At the same time, we do not expect these peaks to carry much of the cosmological information, given the relative increase in cosmic variance, as shown in the spread among the covariance matrix realisations in Fig. 4.1. Nevertheless, this may be improved in future work. We again normalise the counts to a density $n(\nu) = \mathrm{d}N/\mathrm{d}\kappa$, and apply a logarithmic transform $\ln[n(\nu) + \epsilon]$ before emulation. For the height distribution, we use the same 15-component PCA + amplitude GP emulator as described for the peak redshift distribution.

4.A.3 Angular clustering

The angular clustering of the peaks is measured in 12 logarithmically-spaced bins between $\theta = 0.1 - 3.0\,\mathrm{deg}$. We chose this range because it corresponds to the angular range we also use for our $\xi_{+/-}$ 2PCFs. To reduce the dynamic range and improve the emulation accuracy, we transform the correlation function distributions as $\tilde{\omega}(\theta) = \ln[\theta\,\omega(\theta)]$. For $\tilde{\omega}(\theta)$, we adopt a similar PCA GP emulator.

4.A.4 Leave-one-out test

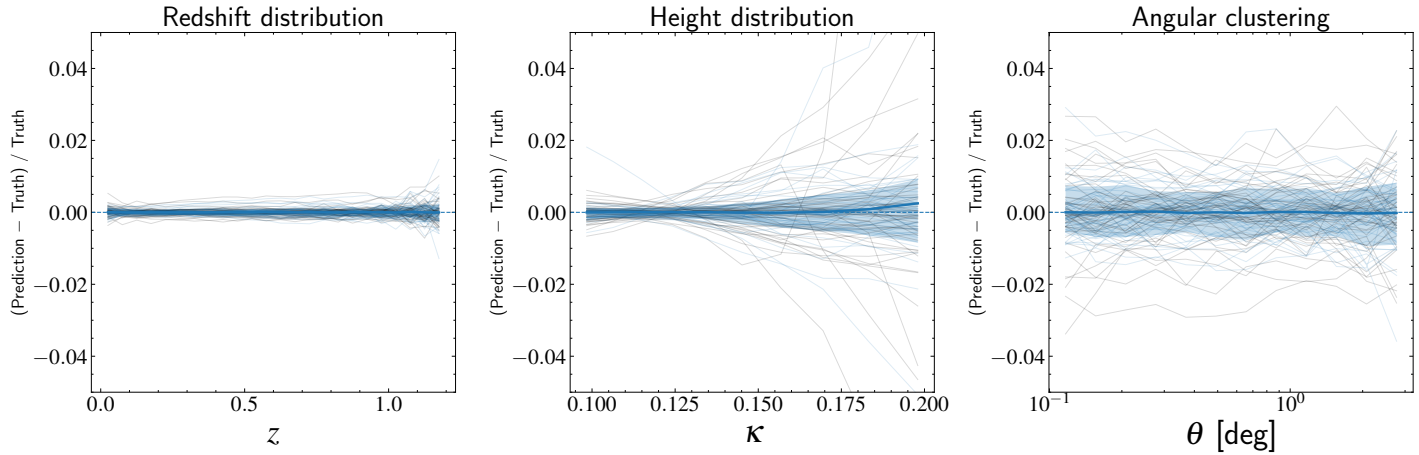


Figure 4.7: Leave-one-out emulator test showing the fractional accuracy of the predicted statistics by training on all but one hypercube node and predicting the statistics at the cosmology of the final hypercube node for the SNR > 3 WL peak redshift distribution (left), height distribution (middle) and angular clustering (right), corresponding to the statistics shown in Fig. 4.1. The emulators show per cent-level accuracy across the complete ranges for which the statistics are used in the forecast.

We assess the emulator accuracy by performing a leave-one-out (LOO) test. Each of the 2×50 training simulations is withheld in turn, and an emulator is trained on the remaining 99 nodes. We then evaluate the prediction at the cosmology of the remaining node and compare it to the actual measurements. For the fiducial forecast setup considered in this work, Fig. 4.7 shows the emulator's fractional accuracy on the log-transformed quantities. The left panel shows the redshift distribution, the central panel the height distribution, and the right panel the angular clustering. For each statistic, averaged over the entire emulated range, the emulator typically has an accuracy of within 1 per cent. The mean 1σ accuracy averaged across all bins is 0.09, 0.40, and 0.58 per cent for the redshift distribution, height distribution, and angular clustering, respectively. Additionally, for the clustering, although the simulations are predicted with per-cent-level accuracy on average, there are individual simulations in which the angular clustering is mispredicted by up to ≈ 3 per cent across all angular scales.

We stress that we designed the emulators to perform well in the fiducial forecast setup, i.e., with 1-arcmin smoothing while inferring all 10 cosmological parameters simultaneously. We did not change the architecture for the 0.5- or 2.0-arcmin smoothing tests. Inspection of the emulator LOO accuracy tests shows that the emulators do not perform much worse at those smoothing scales.

4.B Correlation matrix

In this appendix, we present the correlation matrix for our fiducial forecast setup. Fig. 4.8 shows the Pearson correlation coefficient values, $\rho_{i,j} = C_{ij}/(\sigma_i \sigma_j)$, where C_{ij} is the i,j-th component of the covariance matrix and σ_i and σ_j denote the standard deviation of the i-th and j-th component, respectively, for the forecast corresponding to the black contours in Figs. 4.2 & 4.9. The auto- and cross-correlation blocks corresponding to separate (combinations of) statistics are separated by solid black lines. From bottom left to top right, the figure shows the clustering, $\omega(\theta)$, height distribution, $\mathrm{d}n(\kappa)/\mathrm{d}\kappa$, and redshift distribution, $\mathrm{d}n(z)/\mathrm{d}z$. The x-axis indicates the angular, WL convergence, or redshift bin values. The clustering shows strong correlations, particularly between individual low-angular-separation bins and between individual high-separation bins, whereas the other statistics and cross-correlations show milder correlations, though there is still typically some moderate correlation.

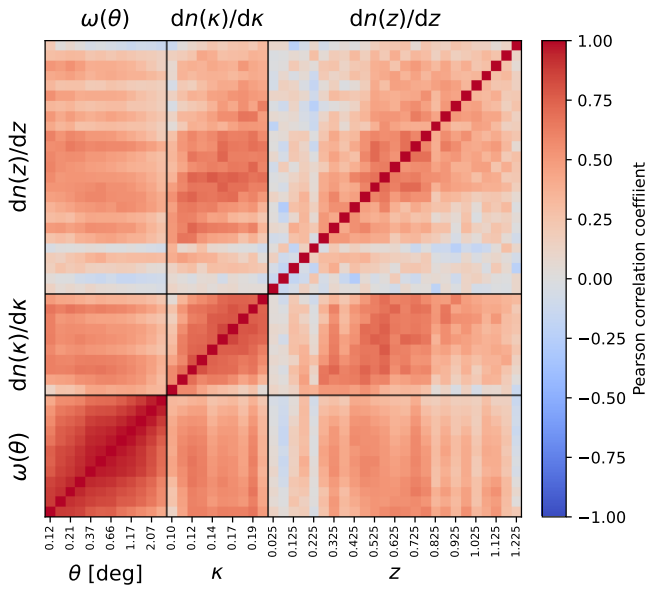


Figure 4.8: Matrix of Pearson correlation coefficients for the inference for the fiducial inference setup. The solid black curves separate the bins corresponding to the different WL peak statistics. Bottom left: angular clustering of SNR > 3 peaks. Centre middle: height distribution. Top right: Redshift distribution of SNR > 3 WL peaks. The different clustering bins are strongly correlated, whereas the other statistics and cross-correlations show weaker correlations.

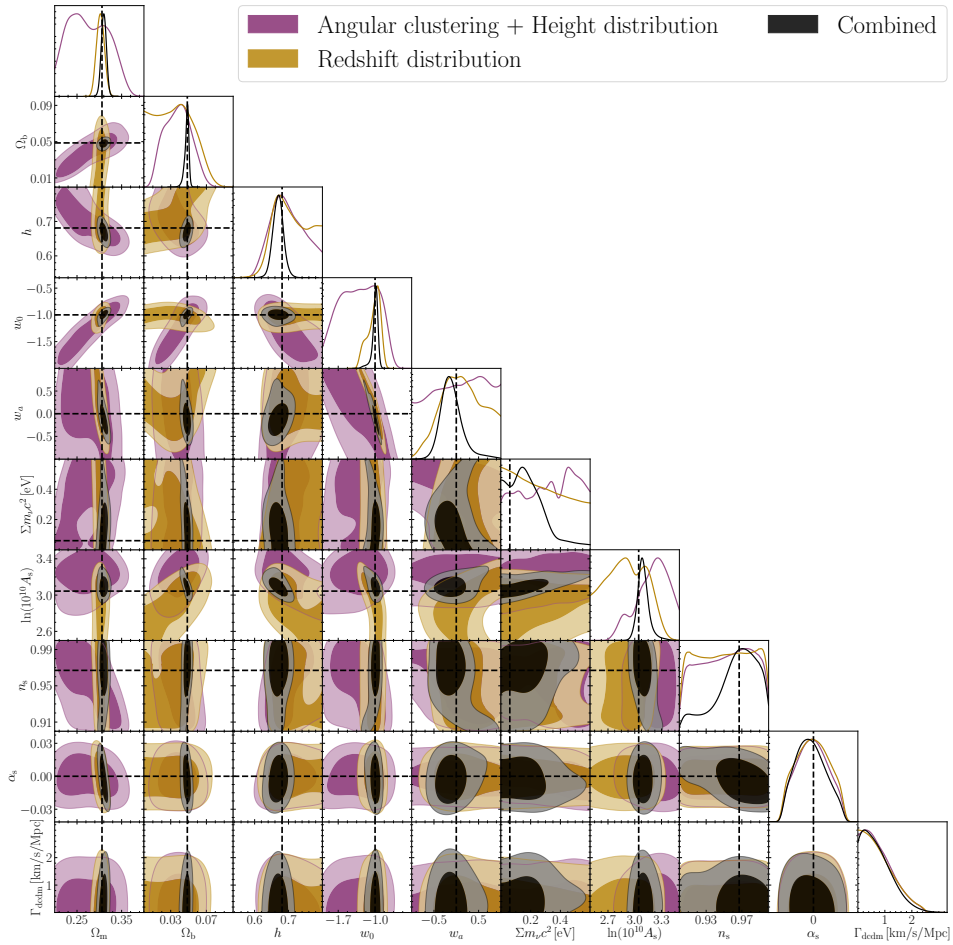


Figure 4.9: Corner plot of the cosmological parameter forecast using the angular clustering and height distributions (purple), redshift distribution (orange) or the combination of the three statistics (black) of WL peaks. The contours indicate the 1- and 2σ credible intervals. Different statistics are sensitive to different parameters, with the redshift distribution being most sensitive to Ω_m, w_0, w_a, and the angular clustering and height distribution jointly probing $\ln(10^{10}A_\mathrm{s})$. Additionally, due to the breaking of degeneracies, the combination of statistics is also able to probe Ω_b and h, whereas $\sum m_\nu c^2$, n_s, α_s, and Γ_dcdm remain unconstrained.

4.C Cosmological constraints from WL peaks

In this appendix, we present the full corner plot of the 2D posteriors for the analysis of the three different WL peak statistics jointly, the redshift distribution individually and the angular clustering and height distribution combined, with 1 ar-

cmin smoothing, whose individual 1D parameter posteriors were already shown in Fig. 4.2. Fig. 4.9 shows the posteriors corresponding to the forecast using the redshift distribution (orange), height distribution and angular clustering (purple) or all three distributions (black).

Individually, the WL peak statistics cannot provide strong constraints over the entire 10D parameter space. Ω_m, w_0, and w_a are most tightly constrained by the redshift distribution. The peak height distribution and angular clustering constrain $\ln(10^{10}A_\mathrm{s})$ best, whereas Ω_b and h are primarily constrained due to breaking of degeneracies in the combined forecast, as can be seen in the $\Omega_\mathrm{m}-\Omega_\mathrm{b}$ and $\Omega_\mathrm{m}-h$ panels. $\Sigma m_\nu c^2, n_\mathrm{s}, \alpha_\mathrm{s}$, and Γ_dcdm remain unconstrained or recover the imposed prior even in the combined case.

The corner plot illustrates the complementarity of combining the different statistics to break degeneracies and obtain tighter constraints. In particular, in the Ω_m–Ω_b plane, no individual statistic can simultaneously constrain both parameters, whereas their combination yields a clearly localised posterior, which seems to result from the combination of the angular clustering and the redshift distribution.

Some of the 2D contours, particularly when parameters are not well constrained, show multimodal posteriors, which may indicate an insufficiently regular emulator. This particularly holds for the angular clustering and height distribution combination, as for instance seen in the $\Omega_\mathrm{m}-\Sigma m_\nu c^2$, where Appendix 4.A already showed that individual nodes can be misestimated on average by several per cent. This behaviour could indicate that improvements in emulation or additional hypercube nodes are necessary to obtain accurate predictions across the entire 10D cosmological parameter space.

4.D Impact of individual cosmological parameters on the WL peak statistics

In this appendix, we show the impact of additional individual cosmological parameters on the WL peak statistics using emulator predictions, analogous to the distributions shown for varying Ω_m in Fig. 4.3. Nine out of the ten parameters that are input to the emulator are set to their mean value, and the tenth parameter is varied within its range within the hypercube. Figs. 4.10, 4.11, 4.12, and 4.13, respectively, show the redshift distribution (left), height distribution (centre) and angular clustering (right) for varying h, w_0, w_a, and Ω_b as indicated by the colour bar. Individually varying all remaining hypercube parameters results in smaller variations than seen for these parameters.

Varying h affects each of the WL peak statistics. The amplitude of the redshift distribution has a positive correlation with h. The height distribution is more sensitive to h, as larger values of h systematically increase the abundance of peaks

at fixed κ, with larger κ values being preferred for larger h values. For the angular clustering distributions, larger h increases the signal for $\theta < 1$ deg.

Stronger evolving DE, with larger values of w_0 or w_a, leads to a suppression of each of the three WL peak statistics. The behaviour of w_0 and w_a is very similar across the three statistics. Only above $z \approx 1$ do larger values of w_0 lead to a small enhancement of the number of peaks, whereas for the w_a values considered, there is only a suppression.

Varying Ω_b only has a minor impact on the redshift distribution and angular clustering, whereas the height distribution shows a stronger dependence, with higher values of Ω_b leading to a suppression of the number of high-valued peaks.

Figure 4.10: Same as Fig. 4.3 but varying h. Larger values of h generally enhance the amplitude of each of the three signals. Additionally, the height distribution is enhanced preferentially at large κ and the clustering signal at small θ.

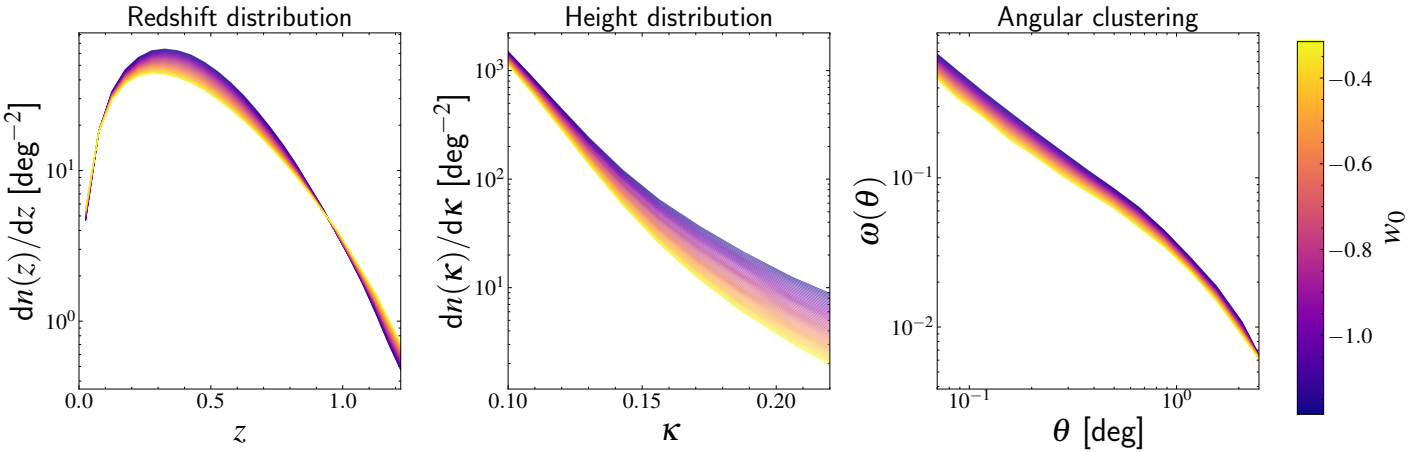


Figure 4.11: Same as Fig. 4.3 but varying w_0. Stronger evolving DE, corresponding to less negative values of w_0, suppresses each of the considered statistics, with the suppression most pronounced at low redshift, high κ and small θ.

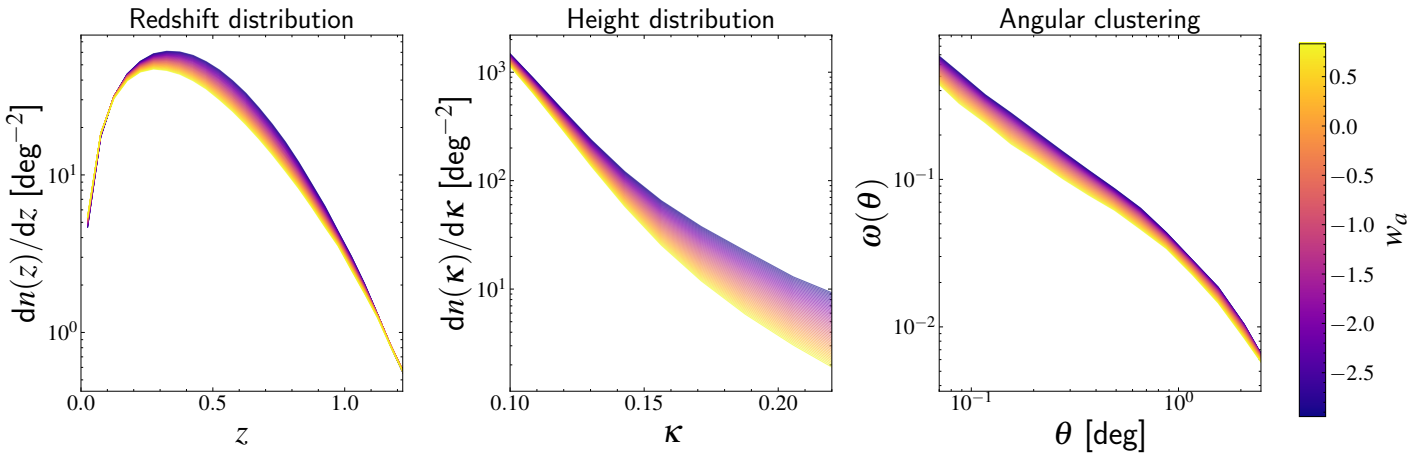


Figure 4.12: Same as Fig. 4.3 but varying w_a. The impact of varying w_a is very similar to that of varying w_0 as shown in Fig. 4.11.

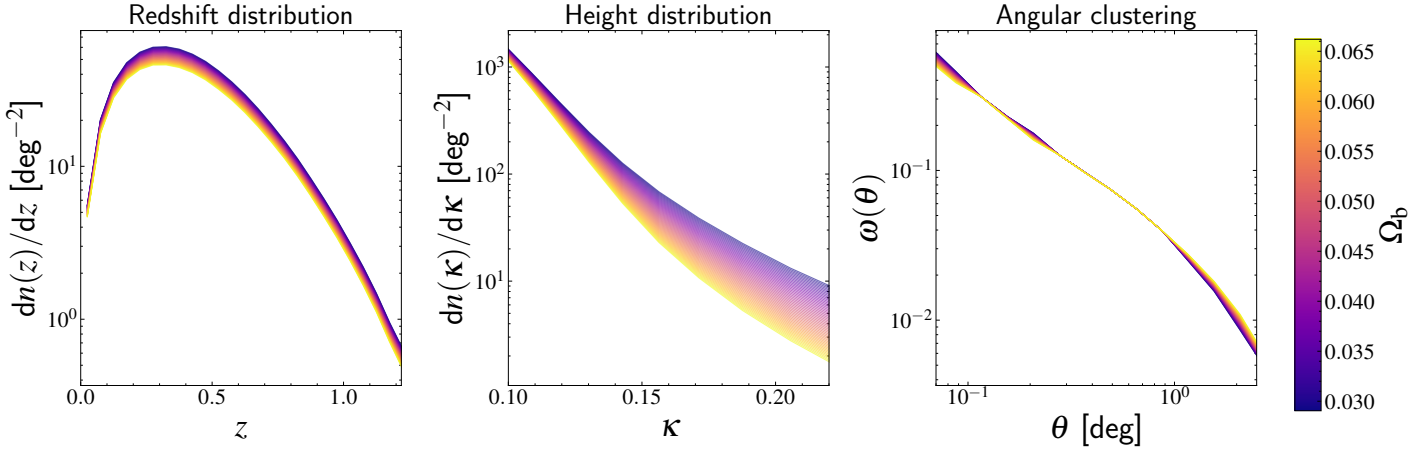


Figure 4.13: Same as Fig. 4.3 but varying Ω_b. The impact of varying Ω_b is most distinct on the height distribution, where lower values of Ω_b lead to more higher-valued WL peaks at fixed κ.

5 First step toward matter power spectrum reconstruction with Stage III weak gravitational lensing surveys

Authors: **Jeger C. Broxterman** and Konrad Kuijken
Published in Astronomy & Astrophysics, 692, A201 (2024)

Abstract

Weak gravitational lensing (WL) surveys provide insight into the matter distribution over an extensive range of scales. Current WL results are in mild tension with cosmic microwave background measurements from the early Universe. Reconstructing the matter power spectrum from their measurements instead of condensing the information into a single cosmological parameter may help locate the origin of these differences. To investigate the cosmic shear measurements of Stage-III WL surveys, we compare their tomographic data by assuming a simple parametric model for the matter power spectrum. The model allows the comparison of surveys with different characteristics and, in an agnostic approach, gives insight into the shape of the matter power spectrum preferred by the data without assuming a cosmological model. For the matter power spectrum, we assume a double power law model in scale factor and wavenumber. The best-fitting amplitude and exponents are inferred in an MCMC analysis. We identify the scales to which the data is most sensitive. We test the sensitivity to different assumptions of the intrinsic alignment strength. We find that Stage-III surveys' constraining power on the power spectrum shape and evolution is still limited. Most information can be summarized as an overall amplitude at a pivot point in wave number and scale factor, while constraints on the power law indices are considerably weaker. Nevertheless, all surveys prefer a weaker rate of growth from $z = 0.5$ to 0.1 than predicted. The assumed intrinsic alignment strength is found to have no significant impact on the measured parameters and goodness of fit. Direct estimates of the matter power spectrum from Stage-III weak lensing surveys can, in principle, be used to locate the physical origin of the observed S_8 tension. We present a simple methodology for the first steps in this direction but find that current constraints are still weak.

5.1 Introduction

The concordance model of cosmology, the Λ cold dark matter (ΛCDM) model, can fit an extensive range of observations with great accuracy (see e.g. Lahav & Liddle, 2022). Examples are the fluctuations of the cosmic microwave background (CMB, Planck Collaboration et al., 2020a, *Planck*), baryon acoustic oscillations (BAO, Adame et al., 2025a), and cosmic shear, the slight distortion of galaxy images by weak gravitational lensing (WL) of the large-scale structure of the Universe (Kilbinger, 2015). With the increasing accuracy of cosmological surveys, several tensions between different probes have emerged. Most notably, the tension in the value of the Hubble constant (H_0) inferred from high-redshift *Planck* CMB measurements or local distance ladder measurements (Riess et al., 2021), and the value of $S_8 \equiv \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$.[1] Here, the value of the *Planck* CMB measurements of the early Universe is in slight tension with the value measured from cosmic shear measurement of the local Universe (Hildebrandt et al., 2017; Asgari et al., 2021; Secco et al., 2022; Dark Energy Survey and Kilo-Degree Survey Collaboration et al., 2023; Li et al., 2023). The main proposed solutions to solving the S_8 tension involve unrecognized systematics or new physics (see e.g. Lucca, 2021; Tanimura et al., 2023; Joseph et al., 2023).

The tension between the probes may have different physical origins. The CMB measures linear scales and the early Universe, whereas WL probes lower redshifts and non-linear scales. Assuming a physical origin, the tension may originate either from differences in scale (high vs. low wavenumber; k) or in time (high vs. low redshift; z).

Amon & Efstathiou (2022) explore the first option, introducing a one-parameter phenomenological model that modifies the power spectrum by fixing it at linear scales to the *Planck* prediction and predicting the non-linear suppression required to reconcile cosmic shear measurements from the Kilo-Degree Survey (KiDS) with *Planck*. They remain agnostic as to what causes the suppression: obvious candidates are misinterpreted baryonic components or new physics. However, the model predicts the suppression of the non-linear power spectrum using only dark matter (DM), and it does not contain any baryonic-informed components. Although baryonic suppression and non-linear collapse manifest themselves in roughly the same wavenumber regime, the suppression they model is on slightly larger scales and has a different shape than predicted by state-of-the-art cosmological hydrodynamical simulation calibrated to primarily X-ray data (see e.g. Fig. 1 of Bigwood et al., 2024; McCarthy et al., 2017; Pakmor et al., 2023; Schaye et al., 2023; Schaller et al., 2025).

[1] Where Ω_m is the matter density parameter, and σ_8 the amplitude of the linear theory matter power spectrum parametrized as the root mean squared mass density fluctuation in spheres of radius 8 h^{-1} Mpc, both at $z = 0$.

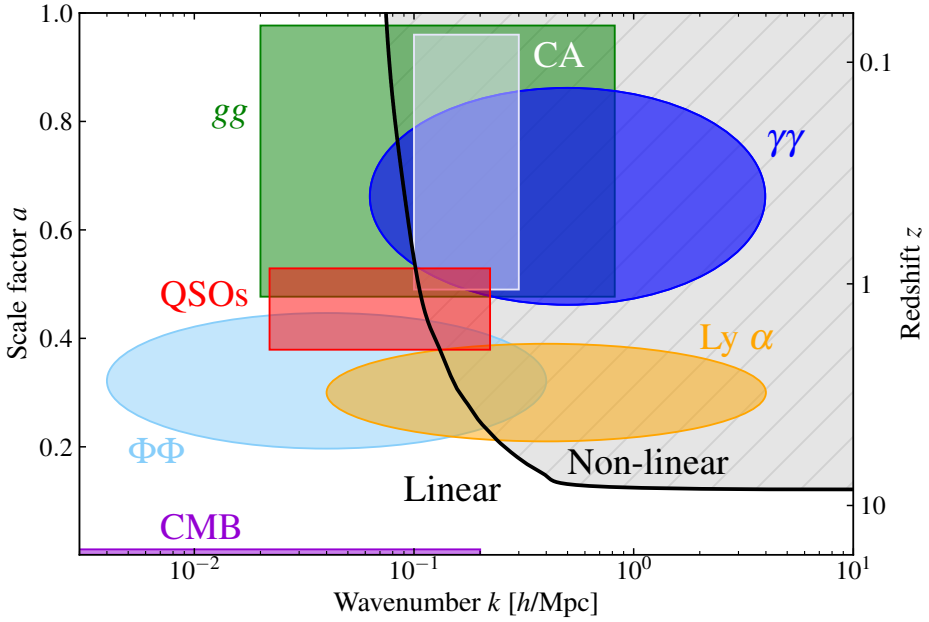


Figure 5.1: Observational cosmological probes of the matter power spectrum and the wavenumber (k), scale factor (a), and redshift (z) ranges to which they are approximately sensitive. From early to late Universe: the cosmic microwave background (CMB, purple), CMB lensing ($\Phi\Phi$, light blue), the Lyman-alpha forest (Ly α, orange), Quasi-Stellar Objects (QSOs, red), galaxy clustering (gg, green), cosmic shear ($\gamma\gamma$, dark blue), and cluster abundance (CA, white). The black curve separates the linear and non-linear collapse regimes. Constraining the matter power spectrum from different probes will allow the localization of discrepancies in the two-dimensional (k, a) plane and provide insight into the physical origin of tensions between the measurements of different probes.

Preston et al. (2023) extended the analysis to include the most recent cosmic shear measurement from the Dark Energy Survey (DES), finding they require a somewhat lower suppression to resolve the tension with *Planck*. However, current measurements are not precise enough to constrain the baryonic feedback models. Recent findings suggest the simulations calibrated to X-ray data, which probe the central regions of haloes, underpredict baryonic feedback as they cannot reproduce Sunyaev-Zeldovich (SZ) measurements, which are more sensitive to the outskirts of haloes (Amodeo et al., 2021; Bigwood et al., 2024; Hadzhiyska et al., 2024). Although it might help resolve the S_8 tension, it raises a different problem because the X-ray and SZ measurements are no longer consistent with the same baryonic feedback model.

In principle, we can distinguish tensions between measurements at different scales and times by directly constraining the matter power spectrum in the wavenumber (k) and scale factor (a) plane. Fig. 5.1 illustrates the sensitivity of different observational cosmological probes in the (k,a) plane. The different probes are cosmic shear ($\gamma\gamma$), galaxy clustering (gg, e.g. Reid et al., 2010; Gil-Marín et al., 2016), CMB, CMB lensing ($\Phi\Phi$, e.g. Planck Collaboration et al., 2020b), cluster abundance (CA, e.g Tegmark et al., 2004), Quasi-Stellar Objects (QSOs, e.g. Adame et al., 2025b), and the Lyman alpha forest (Ly α, e.g. Palanque-Delabrouille et al., 2015; Karaçaylı et al., 2024). The squares correspond to probes of which their emitted radiation is directly observed, whereas the ellipses correspond to processes happening along the line of sight between emission and observation. The black curve separates the linear and non-linear collapse regimes. The estimate is obtained from CLASS (Blas et al., 2011) by determining the wavenumber for which the non-linear estimate differs more than 5% from the linear prediction. The difference between the different cosmic times and scales probed by the CMB and cosmic shear, as described before, is clearly visible in the Figure. Comparing different probes and their cross-correlations allows for the localization of the tension, which should inform about the origin of the observed discrepancies.

Preston et al. (2024) propose constraining the matter power spectrum from cosmic shear surveys. They estimate the constraining power of the matter power spectrum using synthetic cosmic shear measurements for the final data release of Stage-III and Stage-IV WL surveys. They find that while current cosmic shear surveys should only marginally be able to distinguish differences from a ΛCDM dark matter-only (DMO) model, next-generation surveys should provide clear constraints on the shape of the matter power spectrum at (mildly) non-linear scales.

Here, as a first step, we infer the best-fitting shape of the matter power spectrum using the latest cosmic shear measurement of Stage-III WL surveys. We assume a simple parametric model that allows us to quantify the evolution of the matter power spectrum with k and a preferred by current data in an agnostic approach. The model allows the comparison of different surveys with different survey properties and does not assume a cosmological model for the matter power spectrum. We identify the scales to which the data is most sensitive and test the impact of the assumed intrinsic alignment (IA) on the goodness of fit.

In Sect. 5.2, we summarize the relevant WL theory to estimate the measured cosmic shear two-point statistics from the matter power spectrum. Section 5.3 describes the data, and the assumed parametric model is introduced in Sect. 5.4. In Sect. 5.5, we present the measurements of the best-fitting parametric model of the matter power spectrum. We also compare our results to the best-fitting ΛCDM predictions. The main findings are summarized in Sect. 5.6.

5.2 Theory

In this section, we briefly summarize the theoretical framework for cosmic shear measurements as projected statistics of the matter power spectrum.

Cosmic shear measures the systematic distortion of galaxy images from WL by the large-scale structure. The measured signal is comprised of the signal induced by weak gravitational lensing (G), the intrinsic alignment of the galaxies (I), and the cross-correlation of the two effects. The observed cosmic shear angular power spectrum ($C_{\epsilon\epsilon}^{(ij)}$) between two populations of galaxies in different tomographic bins i and j, as a function of multipole moment ℓ, is given by

$$C_{\epsilon\epsilon}^{(ij)}(\ell) = C_{\rm GG}^{(ij)}(\ell) + C_{\rm GI}^{(ij)}(\ell) + C_{\rm II}^{(ij)}(\ell). \tag{5.1}$$

Current cosmic shear surveys use B-mode statistics as null tests, finding no evidence for non-zero B-modes or not using the range where the B-modes are non-zero due to unrecognized systematics. Here, we assume the B-modes are 0 and only use the E-mode signal such that $C_{\epsilon\epsilon}^{(ij)}(\ell) = C_{\epsilon\epsilon,{\rm E}}^{(ij)}(\ell)$. Assuming the modified Limber approximation, the angular power spectra are estimated from the matter power spectrum ($P_{\rm m}$) as (Limber, 1953; LoVerde & Afshordi, 2008)

$$C_{\rm ab}^{(ij)}(\ell) = \int_0^{\chi_{\rm hor}} {\rm d}\chi \, \frac{W_{\rm a}^{(i)}(\chi) W_{\rm b}^{(j)}(\chi)}{f_{\rm K}^2(\chi)} \, P_{\rm m}\left(\frac{\ell + 1/2}{f_{\rm K}(\chi)}, z(\chi)\right), \tag{5.2}$$

where $f_{\rm K}(\chi)$ is the angular diameter distance, χ the comoving distance, z the redshift, and ${\rm a}, {\rm b} \in {\rm I}, {\rm G}$. The integral runs until the comoving line of sight horizon that is the edge of the galaxy survey ($\chi_{\rm hor}$). The WL kernel W is given by (Kaiser, 1992)

$$W_{\rm G}^{(i)}(\chi) = \frac{3H_0^2\Omega_{\rm m}}{2c^2} \frac{f_{\rm K}(\chi)}{a(\chi)} \int_\chi^{\chi_{\rm hor}} {\rm d}\chi' \, n_{\rm s}^{(i)}(\chi') \frac{f_{\rm K}(\chi' - \chi)}{f_{\rm K}(\chi')}, \tag{5.3}$$

where c is the speed of light, $a = 1/(1+z)$ the scale factor, and $n_{\rm s}(\chi)$ is the comoving distance source distribution that relates to the source redshift distribution as $n_{\rm s}(z){\rm d}z = n_{\rm s}(\chi){\rm d}\chi$. For the IA kernel, we assume the non-linear alignment (NLA) model (Bridle & King, 2007)

$$W_{\rm I}^{(i)}(\chi) = -A_{\rm IA} \left(\frac{1+z(\chi)}{1+z_{\rm piv}}\right)^{\eta_{\rm IA}} \frac{C_1 \rho_{\rm cr} \Omega_{\rm m}}{D(a[\chi])} \, n^{(i)}(\chi), \tag{5.4}$$

with $A_{\rm IA}$ the IA amplitude, $z_{\rm piv}$ a finite redshift pivot point, $\eta_{\rm IA}$ the redshift dependence exponent which we set to 0, $C_1\rho_{\rm cr} = 0.0134$, and D the linear growth factor. We assume a flat Universe such that $f_{\rm K}(\chi) = \chi$. We either adopt $A_{\rm IA} = 0$ or 1 to quantify whether including intrinsic alignment better fits the data.

We use three different cosmic shear observables in our analysis. Each observable is a projection of the angular power spectra but is sensitive to different scales

as they depend on different filter functions. The first observable is the shear two-point correlation functions (2PCFs, $\xi_{+/-}$);

$$\xi_{+/-}^{(ij)}(\theta) = \int_0^\infty \frac{\mathrm{d}\ell\ell}{2\pi} \mathcal{J}_{0/4}(\ell\theta) C_{\epsilon\epsilon}^{(ij)}(\ell), \tag{5.5}$$

where $\mathcal{J}_0$ and $\mathcal{J}_4$ are cylindrical Bessel functions of the first kind of the zeroth and fourth order, respectively. Secondly, we consider Complete Orthogonal Sets of E/B-Integrals (COSEBIs). An advantage of the COSEBIs is that close to all cosmological information is contained in the first few n-modes, thus reducing the necessary computational time (Asgari et al., 2012). The COSEBIs (E_n) are estimated using (Schneider et al., 2010)

$$E_n^{(ij)} = \int_0^\infty \frac{\mathrm{d}\ell\, \ell}{2\pi} C_{\epsilon\epsilon}^{(ij)}(\ell)\, W_n(\ell), \tag{5.6}$$

where the weight function, $W_n(\ell)$, is a Hankel transforms given by Equation 9 of Asgari et al. (2021). Similar to Asgari et al. (2021), we use the first 5 COSEBIs n-modes.

Finally, we consider band power spectra ($C_{\mathrm{E},l}$, band powers). This statistic is an angular average of the two-point correlation function that has a weak correlation between values at different wavenumbers and is given by (Schneider et al., 2002)

$$C_{\mathrm{E},l}^{(ij)} = \frac{1}{2\mathcal{N}_l} \int_0^\infty \mathrm{d}\ell\, W_{\mathrm{EE}}^l(\ell) C_{\epsilon\epsilon}^{(ij)}(\ell), \tag{5.7}$$

where the normalisation $\mathcal{N}_l$ and kernels W_{EE} are given by equations 18 and 26 of Joachimi et al. (2021), respectively. We assume the best-fitting ΛCDM cosmology from *Planck* TT,TE,EE+lowE+lensing for the values of the cosmological parameters and to compute the redshift-comoving distance relation and linear growth factor (Planck Collaboration et al., 2020a).

5.3 Data

We use the latest public data releases of the two-point statistics of three Stage-III WL surveys: the Kilo-Degree Survey data release 4 (KiDS-1000; Kuijken et al., 2019), the Dark Energy Survey year 3 (DES year 3; Abbott et al., 2022), and the Hyper Suprime-Cam Year 3 (HSC year 3; Dalal et al., 2023). Collectively, these surveys are referred to as Stage-III WL surveys. We use each survey's summary statistics, source redshift distributions, and full covariance matrix, as described in Asgari et al. (2021); Hildebrandt et al. (2020); Joachimi et al. (2021) (KiDS-1000); Abbott et al. (2022); Myles et al. (2021); Friedrich et al. (2021) (DES year 3); and Shirasaki et al. (2019); Li et al. (2023); Rau et al. (2023) (HSC year 3). We cut the data vectors at the same angular scales as the fiducial survey analyses. In

Table 5.1, we summarize the main characteristics of the surveys and the statistics we use in this analysis. The table lists the sky area covered, effective source number density ($n_{\rm eff}$), standard deviation of the galaxy shape measurement error, and number of tomographic bins. As indicated in the final column, for KiDS-1000, we use the 2PCFs, COSEBIs, and band powers, whereas for HSC and DES year 3, we only use the real space 2PCFs from Li et al. (2023) and Abbott et al. (2022), respectively. The former allows us to quantify the difference between the same measurements from a single survey and the scales that are probed by different statistics, whereas using all statistics allows us to quantify the variations between different surveys.

Survey	Area [deg^2]	$n_{\rm eff}$ [arcmin^{-2}]	σ_e	No. of tomographic bins	statistics used in this analysis
KiDS-1000	1006	6.2	0.26	5	$\xi_{+/-}$, E_n, $\mathcal{C}_{\mathrm{E},l}$
DES year 3	4143	5.9	0.26	4	$\xi_{+/-}$
HSC year 3	416	15	0.24	4	$\xi_{+/-}$

Table 5.1: Survey characteristics. For each survey, the sky area covered; total effective source number density, $n_{\rm eff}$; the standard deviation of the error of galaxy shape measurements, σ_e; the number of tomographic bins; and the statistics considered in this work.

5.4 Methodology

Section 5.2 has provided the formalism to estimate the cosmic shear observables from the matter power spectrum. Here, we provide the model we use in our analysis.

For the matter power spectrum, we assume a double power law in wavenumber (k) and scale factor (a) as

$$P_{\rm m}(k,a) = A\left(\frac{k}{k_{\rm piv}}\right)^{p}\left(\frac{a}{a_{\rm piv}}\right)^{m}, \tag{5.8}$$

where the amplitude A, and exponents p and m are the parameters that will be fitted for. The pivot points $k_{\rm piv}$ and $a_{\rm piv}$ are free as they rescale the amplitude. We will use two sets of pivots: first, we determine the optimal values of the pivot points for each survey and statistic by finding the values of $k_{\rm piv}$ and $a_{\rm piv}$ for which the covariance between the amplitude A and exponents p and m is zero. This method provides the power spectrum amplitude of the point on the (k,a) plane most tightly constrained by the statistic. We will provide the best-fitting parameters of the double power law parameters assuming these pivot points. After, to more straightforwardly compare the three surveys, we adopt a common pivot for all surveys: $k_{\rm piv} = 0.5\,h$/Mpc and $a_{\rm piv} = 0.75$. We estimate the best-fitting parameters that maximize the log-likelihood using the public MCMC sampler EMCEE (Foreman-Mackey et al., 2013). We use flat priors in the range $\log_{10} A \in [0,10]$, $p \in [-2,0.5]$, and $m \in [-5,5]$. We compute the reduced chi-square value between the double power fits and measurements to quantify the goodness of fit.

5.5 Results

5.5.1 Double power law constraints

We first list the results of determining the optimal pivot points for each survey. Table 5.2 lists the best-fitting pivot points in wavenumber and scale factor for each of the 5 statistics considered in this work and for the choice of $A_{\rm IA} = 0$ or 1 in the third and fourth column, respectively. The fifth column states the redshift corresponding to the scale factor pivot point ($z_{\rm piv} = 1/a_{\rm piv} - 1$).

Statistic	$A_{\rm IA}$	$k_{\rm piv}\,[h/{\rm Mpc}]$	$a_{\rm piv}$	$z_{\rm piv}$	$\log_{10} A\,[{\rm Mpc}^3/h^3]$	p	m	$\chi^2_{\rm red}$
KiDS-1000 COSEBIs	0	1.20	0.75	0.34	2.04 ± 0.03	-1.30 ± 0.04	0.3 ± 0.8	1.25
KiDS-1000 COSEBIs	1	1.24	0.75	0.34	2.03 ± 0.03	-1.30 ± 0.04	0.2 ± 0.9	1.26
KiDS-1000 band powers	0	1.03	0.74	0.34	2.16 ± 0.03	-1.25 ± 0.06	-0.1 ± 1.0	1.39
KiDS-1000 band powers	1	0.93	0.74	0.35	2.21 ± 0.03	-1.26 ± 0.06	-0.1 ± 1.1	1.40
KiDS-1000 2PCFs	0	3.64	0.74	0.35	1.44 ± 0.04	-1.24 ± 0.03	0.0 ± 0.8	1.25
KiDS-1000 2PCFs	1	4.75	0.74	0.35	1.30 ± 0.04	-1.24 ± 0.03	-0.3 ± 0.8	1.25
DES year 3 2PCFs	0	2.17	0.79	0.26	1.92 ± 0.03	-1.14 ± 0.02	1.0 ± 0.5	1.14
DES year 3 2PCFs	1	2.01	0.79	0.26	1.97 ± 0.03	-1.14 ± 0.02	1.0 ± 0.5	1.14
HSC year 3 2PCFs	0	0.80	0.71	0.41	2.36 ± 0.03	-1.27 ± 0.03	-0.2 ± 0.4	1.40
HSC year3 2PCFs	1	0.78	0.71	0.41	2.28 ± 0.03	-1.27 ± 0.03	-0.3 ± 0.4	1.38

Table 5.2: Pivot points and best-fitting parameters for the double power law model. For each statistic and choice of intrinsic alignment amplitude ($A_{\rm IA}$) the values of the pivot points that minimize the covariance between the amplitude $\log_{10} A$, and exponents m and p in wavenumber ($k_{\rm piv}$), scale factor ($a_{\rm piv}$), and corresponding redshift ($z_{\rm piv} = 1/a_{\rm piv} - 1$) and the best-fitting parameters for the double power-law parameters; amplitude, $\log_{10} A$; wavenumber exponent, p; scale factor exponent, m. The final column indicates the reduced chi-square value of the double power law fit, $\chi^2_{\rm red}$.

The values show that the HSC survey is the deepest, as the scale factor pivot point corresponds to $z \approx 0.4$. Then, the three KiDS-1000 statistics, which have almost identical depths, probe slightly less deep. Finally, the DES year 3 2PCFs probe the least deep. The HSC 2PCFs probe the largest scales, whereas the KiDS-1000 2PCFs probe the smallest scales.

Table 5.2 also includes the best-fitting parameter values and their 16th and 84th percentiles of the double power law fit. For each fit, the parameter best-fitting values correspond to an inference assuming the listed pivot point for that statistic and the value of intrinsic alignment amplitude. The best-fitting value of the amplitude parameter may, therefore, differ as this is the value of the double power-law model at $(k, a) = (k_{\rm piv}, a_{\rm piv})$. In Appendix 5.A, we show an example of a double fit power law prediction to the KiDS-1000 COSEBIs.

The 3 KiDS-1000 statistics generally find the same values for the power law exponents m and p, which are consistent with the HSC year 3 2PCFs estimates. The DES year 3 2PCFs prefer a weaker dependence on wavenumber and a stronger dependence on scale factor.

Focussing first on the growth rate, quantified by the m exponent, we see that the constraints from WL are weak. Most statistics are consistent with no growth, with only DES year 3 2PCFs preferring a positive growth rate of $m = 1.0 \pm 0.5$ irrespective of the choice of $A_{\rm IA}$.

Stage-III WL surveys primarily probe the Λ-dominated era, during which linear growth eventually comes to a halt. However, this is a gradual process, and in addition, non-linear growth continues for longer: predictions obtained with CLASS (Blas et al., 2011) indicate an increase in the power spectrum between redshift 0.5 and 0.1 of a factor of 1.5 (1.9) at wavenumber $k = 0.1$ (1.0) h/Mpc, corresponding to m values of 1.3 (2.1). Therefore, all data sets we analyze here mildly favor a slower growth than predicted by ΛCDM. This is most apparent in the HSC year 3 2PCFs, which prefer $m = -0.3 \pm 0.4$.

Comparing the three different KiDS-1000 estimates for p, we find the 2PCFs provide the tightest constraint. The 1σ uncertainty on p is a factor two smaller than the uncertainty provided by the band powers estimate. This is not unexpected: for the chosen configuration, the 2PCFs probe the largest range of scales, as illustrated in Fig. 1 of Asgari et al. (2021).

Interestingly, the assumed value for $A_{\rm IA}$ does not markedly affect the results. There are no significant differences between the assumed value for $A_{\rm IA}$. The results using $A_{\rm IA} = 0$ or 1 are always consistent, and including the intrinsic alignment does not give a better fit, as indicated by the reduced chi-square value in the final column. The NLA model assumes that density perturbations are characterized by the Poisson equation. This assumption breaks down under non-linear gravitational collapse, and the model, therefore, cannot physically capture the IA signal over the entire range of scales considered in this work (Krause et al., 2016; Asgari et al., 2021). However, as the current constraining power is still limited

and the values we adopted for the IA strength correspond to those measured by current surveys, we expect that using a different model, such as the Tidal Alignment and Tidal Torquing model (Blazek et al., 2019), or including a redshift (e.g. Joachimi et al., 2011) or luminosity (e.g. Chisari et al., 2016) dependence will not change the insensitivity to IA strength.

The values of the reduced chi-square statistics are reasonable for the model's simplicity. For example, the model does not capture the BAO wiggles and can still fit the data over a broad wavenumber and redshift range. The reduced chi-square values for the KiDS-1000 COSEBIs are 1.25 ($A_{\rm IA} = 0$) and 1.26 ($A_{\rm IA} = 1$). These values indicate the fits are of equal quality as the best-fitting ΛCDM prediction from Asgari et al. (2021), who quote a value of $\chi^2_{\rm red} = 1.2$. The same holds for the KiDS-1000 band powers and 2PCFs.

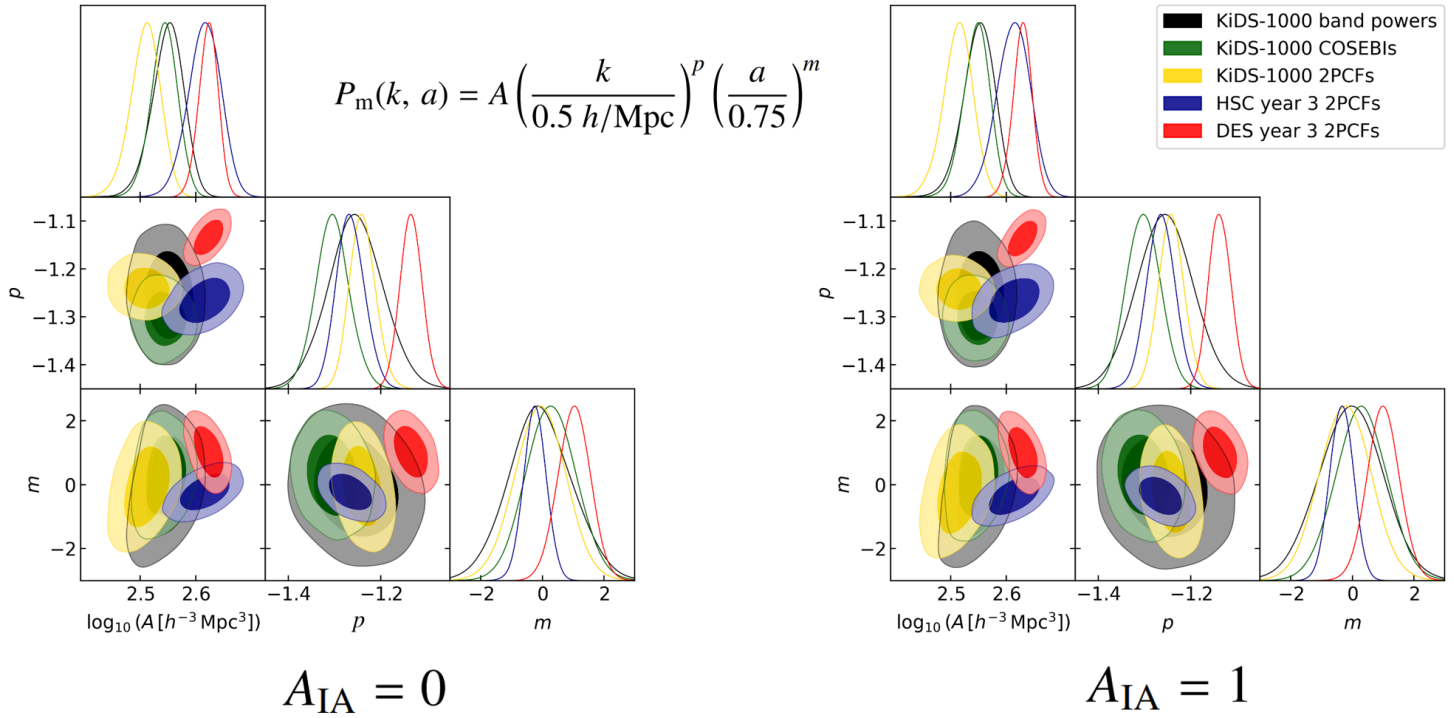


Figure 5.2: Posterior contours for a double power law model as indicated in the top. The results correspond to setting the pivot points to $k_{\rm piv} = 0.5\ h/{\rm Mpc}$ and $a_{\rm piv} = 0.75$. Left: Assuming no contribution from intrinsic alignment. Right: contours from assuming an intrinsic alignment strength set by $A_{\rm IA} = 1$. The panels show the posterior distribution of the inferences using the KiDS-1000 band powers (black), COSEBIs (green), 2PCFs (yellow), HSC year 3 2PCFs (blue), and DES year 3 2PCFs (red). The results are consistent at the $\sim 2\sigma$ level.

In Fig. 5.2, we show the best-fitting parameter constraints assuming a common pivot point for each survey. As the pivot point does not minimize the covariance between the different model parameters, clear correlations are visible between the amplitude parameter and p and m exponents. Still, assuming the same pivot points allows for a direct comparison of the amplitudes measured by the different statistics at the same redshift and wavenumber. The left panel shows the results assuming an intrinsic alignment amplitude of 0 whereas the for the contours obtained in the right panel $A_{\rm IA} = 1$. All KiDS-1000 and HSC year 3 results are consistent at the 2σ level. Comparing the 1D amplitude $\log_{10} A$ posteriors, we see DES and HSC year 3 predict more power than the KiDS-1000 statistics at the 2σ level.

5.5.2 Comparison to ΛCDM predictions

Next, in Fig. 5.3, we compare our power law prediction to the best-fitting ΛCDM predictions from the KiDS-1000 COSEBIs inference assuming $A_{\rm IA} = 1$. The green curve shows the double power prediction. The thin solid curves are all matter power spectrum predictions from the KiDS-1000 COSEBIs ΛCDM inference chains generated using CLASS (Blas et al., 2011) and HMCODE (Mead et al., 2015). These curves are colored by their value of S_8, as indicated by the color bar. The solid and dashed gray curves show the best-fitting and 16th and 84th percentiles of the ΛCDM prediction from the KiDS-1000 COSEBIs (Asgari et al., 2021), respectively. The double power law fit is consistent with the ΛCDM prediction over two decades in wavenumber. At the largest wavenumbers ($k > 0.5$ h/Mpc), the double power law overestimates the power compared to the ΛCDM prediction as well as the largest scales ($k < 10^{-2}$ h/Mpc, not shown), which are beyond the scales to which cosmic shear is most sensitive. Within this range, the simplistic double power law is consistent with the KiDS-1000 COSEBIs ΛCDM prediction. The 16th and 84th percentile range typically covers half a dex, illustrating the matter power spectrum is still poorly constrained.

5.5.3 Suppression of the matter power spectrum

Finally, we show the ratio of the matter power spectrum to a Planck DMO matter power spectrum as a function of wavenumber at $z = 0.35$ in Fig. 5.4. The dashed green curve shows the best-fitting double power law estimate from this work. The solid gray curves and shaded area show the best-fitting and 16th and 84th percentiles of the KiDS-1000 COSEBIs ΛCDM chains, respectively. The dotted red curve shows the prediction assuming the one-parameter parametric model from Amon & Efstathiou (2022) and setting $A_{\rm mod} = 0.69$, which is the value found by Amon & Efstathiou (2022) to reconcile KiDS-1000 with *Planck*.

The simple parametric model from Amon & Efstathiou (2022), our double power law, and the DMO power spectrum are all within the 16th and 84th per-

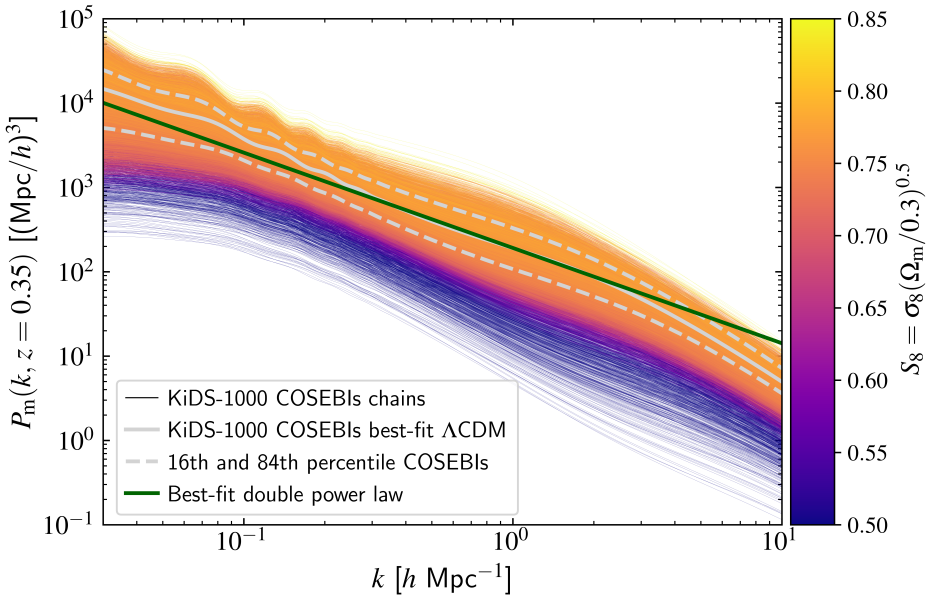


Figure 5.3: Matter power spectrum at $z = 0.35$. The colored curves show the predictions from the KiDS-1000 COSEBIs chains from Asgari et al. (2021), and they are colored by their value of S_8 as indicated by the color bar. The solid and dashed gray curves show the best-fitting and 16th and 84 percentiles of the KiDS-1000 inference, respectively. The green solid curve shows the best-fitting prediction of the $A_{IA} = 1$ double power-law model assumed in this work.

centile prediction from the KiDS-1000 COSEBIs, which reconfirms that current data are not able to constrain detailed deviations from ΛCDM predictions or baryonic physics, as previously predicted in Preston et al. (2024).

5.6 Conclusions

In this letter, we have taken a first step towards reconstructing the shape of the matter power spectrum measured with Stage-III WL lensing surveys. Reconstructing the matter power spectrum from different surveys that are sensitive to different redshift ranges and scales can give insight into the physical origins of observed tensions in cosmological parameters. We do not assume a cosmological model for the shape of the matter power spectrum but instead remain agnostic to its shape by assuming a double power law in wavenumber and scale factor (eq. 5.8). This allows us to directly compare the results of different surveys and quantify the shape of the matter power spectrum preferred by the data. We identify the scales to which the data is most sensitive, and we test different assump-

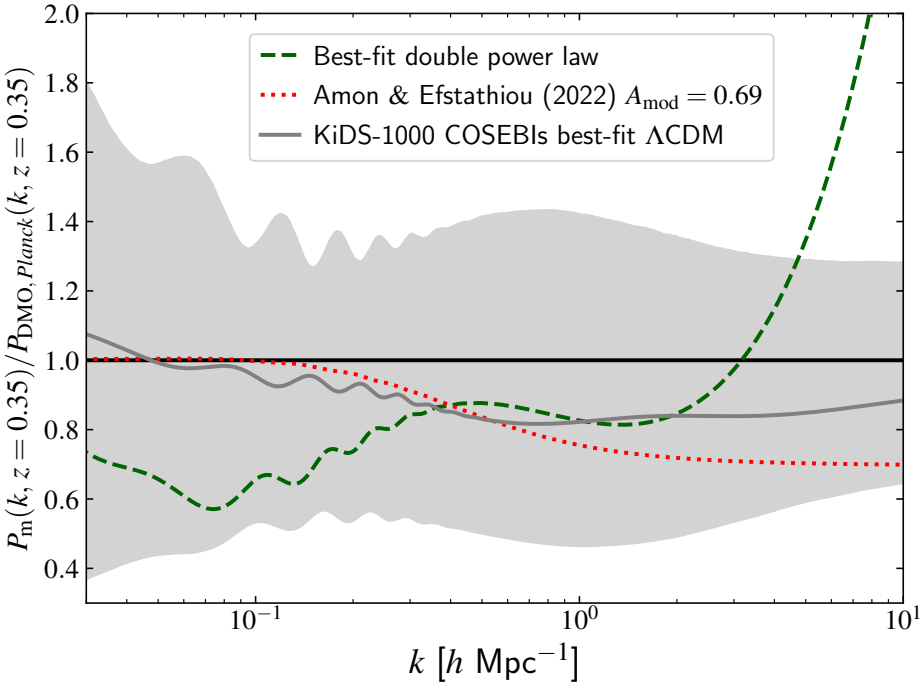


Figure 5.4: Suppression of the matter power spectrum compared to a DMO prediction at $z = 0.35$. The dashed green curve corresponds to the best-fitting prediction from the double power to the KiDS COSEBIs with $A_{\rm IA} = 1$. The solid gray curve shows the best fit of the KiDS-1000 ΛCDM inference from Asgari et al. (2021), and the shaded area covers the 16th to 84th percentile region. The dotted red curve uses the parametric model from Amon & Efstathiou (2022) with $A_{\rm mod} = 0.69$.

tions for the strength of the intrinsic alignment signal. The results highlight the potential of directly constraining the matter power spectrum to explore tensions in current data sets. We find

1. Current cosmic shear surveys do not provide tight enough constraints on the shape of the matter power spectrum to rule out deviations from ΛCDM or realistic baryonic feedback models as the results are still consistent with DMO predictions.

2. Stage-III WL surveys can only poorly constrain the shape and evolution of the matter power spectrum in the low-redshift Universe. Still, they prefer a weaker evolution in the redshift range of 0.5 - 0.1 than predicted by ΛCDM.

3. Including intrinsic alignment of galaxies does not significantly impact the fitting accuracy or best-fitting parameters.

More accurate measurement by next-generation WL surveys carried out by *Euclid* (Euclid Collaboration: Mellier et al., 2025), *Roman* (Spergel et al., 2015), and *Rubin* (LSST Science Collaboration et al., 2009) as well as more accurate modeling will help understand the physical origin of the tensions between different observational cosmological probes.

References

Abbott T. M. C., et al., 2022, Phys. Rev. D, 105, 023520
Adame A. G., et al., 2025a, J. Cosmology Astropart. Phys., 2025, 021
Adame A. G., et al., 2025b, J. Cosmology Astropart. Phys., 2025, 012
Amodeo S., et al., 2021, Phys. Rev. D, 103, 063514
Amon A., Efstathiou G., 2022, MNRAS, 516, 5355
Asgari M., Schneider P., Simon P., 2012, A&A, 542, A122
Asgari M., et al., 2021, A&A, 645, A104
Bigwood L., et al., 2024, MNRAS, 534, 655
Blas D., Lesgourgues J., Tram T., 2011, J. Cosmology Astropart. Phys., 2011, 034
Blazek J. A., MacCrann N., Troxel M. A., Fang X., 2019, Phys. Rev. D, 100, 103506
Bridle S., King L., 2007, New Journal of Physics, 9, 444
Chisari N., et al., 2016, MNRAS, 461, 2702
Dalal R., et al., 2023, Phys. Rev. D, 108, 123519
Dark Energy Survey and Kilo-Degree Survey Collaboration et al., 2023, The Open Journal of Astrophysics, 6, 36
Euclid Collaboration: Mellier Y., Abdurro'uf Acevedo Barroso J., et al., 2025, A&A, 697, A1
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Friedrich O., et al., 2021, MNRAS, 508, 3125
Gil-Marín H., et al., 2016, MNRAS, 460, 4188
Hadzhiyska B., et al., 2024, arXiv e-prints, p. arXiv:2407.07152
Hildebrandt H., et al., 2017, MNRAS, 465, 1454
Hildebrandt H., et al., 2020, A&A, 633, A69
Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, A&A, 527, A26
Joachimi B., et al., 2021, A&A, 646, A129
Joseph M., Aloni D., Schmaltz M., Sivarajan E. N., Weiner N., 2023, Phys. Rev. D, 108, 023520
Kaiser N., 1992, ApJ, 388, 272
Karaçaylı N. G., et al., 2024, MNRAS, 528, 3941
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Krause E., Eifler T., Blazek J., 2016, MNRAS, 456, 207
Kuijken K., et al., 2019, A&A, 625, A2
LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201
Lahav O., Liddle A. R., 2022, arXiv e-prints, p. arXiv:2201.08666
Li X., et al., 2023, Phys. Rev. D, 108, 123518
Limber D. N., 1953, ApJ, p. 134
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Lucca M., 2021, Physics of the Dark Universe, 34, 100899
McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936
Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, MNRAS, 454, 1958
Myles J., et al., 2021, MNRAS, 505, 4249
Pakmor R., et al., 2023, MNRAS, 524, 2539
Palanque-Delabrouille N., et al., 2015, J. Cosmology Astropart. Phys., 2015, 011
Planck Collaboration et al., 2020a, A&A, 641, A6

Planck Collaboration et al., 2020b, A&A, 641, A8
Preston C., Amon A., Efstathiou G., 2023, MNRAS, 525, 5554
Preston C., Amon A., Efstathiou G., 2024, MNRAS, 533, 621
Rau M. M., et al., 2023, MNRAS, 524, 5109
Reid B. A., et al., 2010, MNRAS, 404, 60
Riess A. G., Casertano S., Yuan W., Bowers J. B., Macri L., Zinn J. C., Scolnic D., 2021, ApJ, 908, L6
Schaller M., Schaye J., Kugel R., Broxterman J. C., van Daalen M. P., 2025, MNRAS, 539, 1337
Schaye J., et al., 2023, MNRAS, 526, 4978
Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002, A&A, 396, 1
Schneider P., Eifler T., Krause E., 2010, A&A, 520, A116
Secco L. F., et al., 2022, Phys. Rev. D, 105, 023515
Shirasaki M., Hamana T., Takada M., Takahashi R., Miyatake H., 2019, MNRAS, 486, 52
Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757
Tanimura H., Douspis M., Aghanim N., Kuruvilla J., 2023, A&A, 674, A222
Tegmark M., et al., 2004, Phys. Rev. D, 69, 103501

Appendix

5.A KiDS-1000 COSEBIs prediction

This appendix shows the best-fitting prediction for the double power law model to the KiDS-1000 COSEBIs measurements. The scatter points in Fig. 5.5 are the KiDS-1000 measurements (Asgari et al., 2021), and the red and blue curves are the best-fitting double power law fits with $A_{\rm IA} = 0$ and $A_{\rm IA} = 1$, respectively. The subpanels show the measured E_n as a function of the COSEBI n-modes for the combination of two tomographic bins, as indicated in the upper-right corner of each panel. There is no difference in the fitting accuracy with or without including IA, as seen from the reduced chi-square values in Table 5.2. The double power law estimate fits the same ranges well as the best-fitting ΛCDM prediction from Asgari et al. (2021). For example, the ΛCDM and our double power law estimate both provide good fits to most of the data, but they underpredict the signal for the high-n modes in the $(1-5)$ tomographic bin combination and across the entire range for the $(2-2)$ combination.

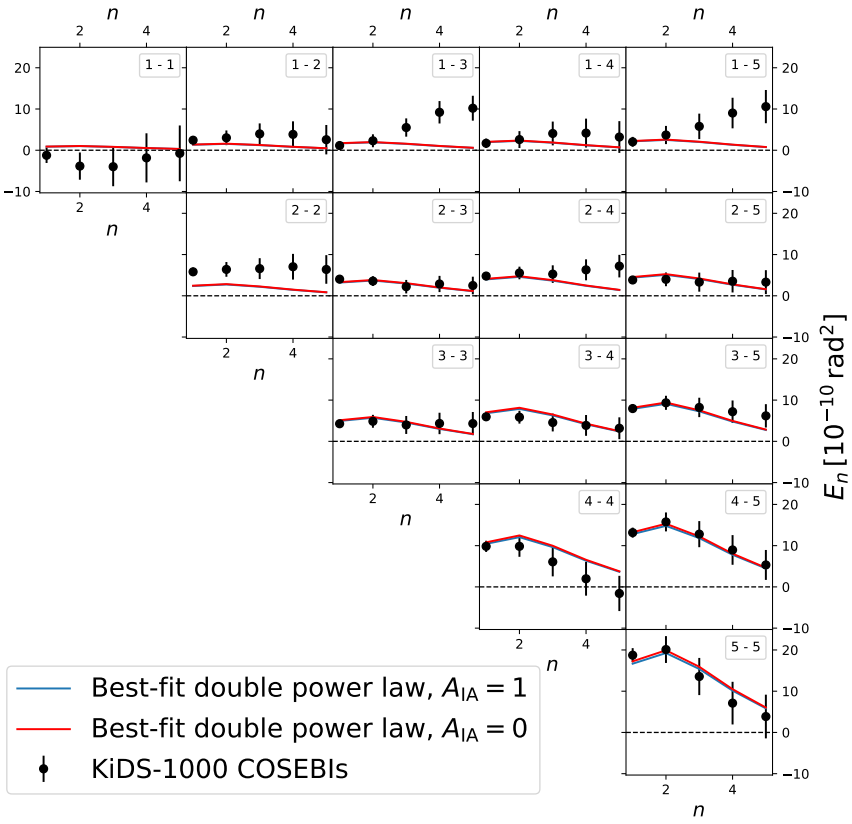


Figure 5.5: KiDS-1000 COSEBIs measurement (black scatter points) with best-fitting $A_{\mathrm{IA}} = 1$ and 0 double power law (solid blue and curves, respectively) for the first 5 COSEBIs n-modes. Each sub-panel corresponds to a combination of two tomographic bins, as indicated in the upper-right corner.

6 Matter power spectrum reconstruction with KiDS-Legacy: Improved internal ΛCDM consistency and preference for strong baryonic feedback

Authors: **Jeger C. Broxterman**, Patrick Simon, Lucas Porth, Konrad Kuijken, Angus H. Wright, Marika Asgari, Maciej Bilicki, Catherine Heymans, Hendrik Hildebrandt, Henk Hoekstra, Benjamin Joachimi, Shun-Sheng Li, Matteo Maturi, Lauro Moscardini, Mario Radovich, Robert Reischke, and Maximilian Von Wietersheim-Kramsta

Published in Astronomy & Astrophysics, 703, L3 (2025)

Abstract

Direct measurements of the matter power spectrum, $P_{\rm m}(k,z)$, provide a powerful tool for investigating the observed tensions between models of structure growth, while also testing the internal consistency of cosmological probes. We analysed the cosmic shear data from the final data release of the Kilo-Degree Survey (KiDS), and present a deprojected $P_{\rm m}(k,z)$, measured in up to three redshift bins. Compared to analyses using previous KiDS releases, we find improved internal consistency in the $z \lesssim 0.7$ regime. At large scales, $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$, our power spectrum reconstruction aligns with ΛCDM predictions with a density fluctuation amplitude $\sigma_8 = 0.81$. Furthermore, at small scales, $k = 3$–$20\,h\,\mathrm{Mpc}^{-1}$, the average matter power spectrum is suppressed by $30\% \pm 10\%\,(\mathrm{stat.}) \pm 4\%\,(\mathrm{sys.})$ with 2.8σ significance relative to a dark matter-only model, consistent with expectations of strong baryonic feedback.

6.1 Introduction

The growth of cosmic structure is quantified by the matter power spectrum, $P_{\rm m}(k,z)$, which is probed across different wavenumbers, k, and redshifts, z, by cosmological observables such as cosmic microwave background lensing (CMB; e.g. Lewis & Challinor, 2006), galaxy clustering (e.g. Reid et al., 2010), the Lyman-alpha forest (e.g. Rauch, 1998), and weak gravitational lensing (WL; e.g. Kilbinger, 2015).

To identify the origins of discrepancies in cosmology inference between different surveys and probes operating at different scales or times, the shape and amplitude of $P_{\rm m}(k,z)$ can be inferred instead of condensing observations by physical model parameters (Amon & Efstathiou, 2022). This approach identifies possible suppression signals relative to a power spectrum in a minimalistic ΛCDM scenario that accounts for structure growth by dark matter-only (DMO) models. The isolated suppression signal may then be used to differentiate signatures from sources within known model uncertainties, such as baryonic effects or new physics (e.g. Preston et al., 2024, 2025). In this work we focus on the interpretation of weak lensing data, for which the uncertainty on the shape of $P_{\rm m}(k,z)$ is substantial. Thus, a functional form is typically assumed for its shape and/or redshift evolution. Proposed models include either perturbative expansions (Ye et al., 2025) or phenomenological models (Seljak, 1998; Pen et al., 2003; Perez Sarmiento et al., 2025).

For this paper we combined the approaches of Broxterman & Kuijken (2024) (hereafter BK24), Simon (2012), and Simon et al. (2025) (hereafter SPBK25) by adopting two different functional forms to extract direct measurements of the matter power spectrum from the cosmic shear data. We compare a rigid double power-law model for $P_{\rm m}(k,z)$ (BK24), with a flexible deprojected regularised model that infers deviations from a best-fitting DMO reference power spectrum (SPBK25). Both methods, applied to the fourth Kilo-Degree Survey data release (Kuijken et al., 2019, hereafter KiDS-1000), show evidence of $P_{\rm m}(k,z)$ inconsistent with the ΛCDM constraints derived from the same data, specifically in the redshift evolution. SPBK25 discuss errors in the adopted redshift distributions within tomographic bins or the modelling of intrinsic alignment (IA) of sources as possible reasons for this inconsistency. Furthermore, the recent work from Doux & Karwal (2025) reports differences between Stage III lensing surveys and *Planck* for $P_{\rm m}(k,z)$ that are in line with those found in BK24 and SPBK25.

The recently published final KiDS data release (KiDS-Legacy; Wright et al., 2024), featuring more area, an improved redshift calibration, one additional higher-redshift tomographic bin, and a new shear catalogue, supersedes KiDS-1000. While the KiDS-1000 results show a mild tension (2σ–3.5σ) with *Planck* CMB measurements (Hildebrandt et al., 2017; Asgari et al., 2021), pointing towards a lower S_8-amplitude of $P_{\rm m}(k,z)$, this tension is now alleviated in the

final cosmic shear analysis (Stölzner et al., 2025; Wright et al., 2025b). This shift in findings motivates us to repeat our previous $P_{\rm m}(k,z)$ analysis using the new higher-quality data.

This Letter presents the result of these KiDS-Legacy $P_{\rm m}(k,z)$ reconstructions with the BK24 and SPBK25 approaches. Section 6.2 summarises the data and describes minor methodological changes relative to the previous works. Section 6.3 presents the deprojected matter power spectrum and ratios relative to the DMO reference for single or multiple redshift bins and compares the results to predictions from state-of-the-art cosmological hydrodynamical simulations. The main results are summarised in Sect. 6.4. Throughout, distances and wavenumbers are reported in comoving units for the Hubble constant $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$.

6.2 Data and analysis

We analysed the KiDS-Legacy shear two-point correlation functions, $\xi_\pm^{(ij)}(\theta)$, between tomographic bins i and j within the angular range $\theta \in [2', 300']$. This analysis incorporated six source redshift distributions, $n_{\rm s}^{(i)}(z)$ with $\int_0^\infty {\rm d}z\, n_{\rm s}^{(i)}(z) = 1$, and the theoretical covariance matrix of $\xi_\pm^{(ij)}(\theta)$ uncertainties, based on Reischke et al. (2025). The methodologies employed in this paper are detailed in BK24 and SPBK25. Here we only discuss specific changes to the original set-up and the data used. Appendix 6.B provides additional details on the original methods. Compared to KiDS-1000, major improvements include updated angular scale cuts, new tomographic bins including an additional sixth tomographic bin extending to a photo-z regime of $1.14 < z_{\rm B} \leq 2$ with deeper i-band imaging, a larger survey area (1347 deg^2, a $\sim 34\%$ extension), improved redshift distribution calibration methods (Wright et al., 2025a) leveraging a larger calibration sample (five times the number of spectroscopic redshifts), new multiband image simulations for shear calibration (Li et al., 2023), and a new IA model featuring redshift evolution. The combined increase in survey area and the extra tomographic bin leads to a roughly 3.5-fold increase in the probed cosmic volume.

The novel IA model is motivated by the observed dichotomy between red and blue galaxies, and the observed scaling of the IA strength with halo mass (e.g. Fortuna et al. 2025, Navarro-Gironés et al. 2025). Consistent with the KiDS-Legacy cosmological analysis, we therefore updated the IA kernel to

$$W_{\rm I}^{(i)}(\chi) = \underbrace{-A_{\rm IA}\, f_{\rm r}^{(i)} \left(\frac{\langle M_{\rm h}\rangle^{(i)}}{M_{\rm h,pivot}}\right)^{\beta} \frac{C_1\,\rho_{\rm cr}\,\Omega_{\rm m}}{D_+(\chi)}}_{=:F^{(i)}(\chi)} n_s^{(i)}(z[\chi])\,\frac{{\rm d}z(\chi)}{{\rm d}\chi}\,, \tag{6.1}$$

where χ is the comoving distance, $z(\chi)$ is the redshift at χ, $A_{\rm IA}$ is the IA amplitude, $\Omega_{\rm m}$ is the total matter density relative to the critical density, $D_+(\chi)$ is the linear growth factor (defined such that $D_+(0) = 1$), $C_1\,\rho_{\rm cr} = 0.0134$ is a constant,

$M_{\rm h,pivot} = 10^{13.5}\,h^{-1}\,M_\odot$ is the pivot halo mass, $\langle M_{\rm h}\rangle^{(i)}$ is the average halo mass in z-bin i, and β is the power-law exponent denoting the evolution with halo mass. Our fractions of red galaxies, $f_{\rm r}^{(i)}$, use the values in Table B.1 of Wright et al. (2025b).

In our IA upgrade, the deprojected regularisation technique now employs $F^{(i)}(\chi)$ for bin i, instead of the previous $F(\chi)$, rendering the IA amplitude z-dependent. Furthermore, we marginalised the posterior of $P_{\rm m}(k,z)$ over the distribution of $A_{\rm IA}$, β, and $\langle M_{\rm h}\rangle^{(i)}$ from the KiDS-Legacy cosmological analysis (Wright et al., 2025b). This was done after asserting the Gaussian prior in SPBK25 of $\Omega_{\rm m} \sim \mathcal{N}(0.305, 0.012)$ for the angular diameter distance in a flat ΛCDM universe. The DMO reference power spectrum, $P_{\rm m}^{\rm DMO}(k,z)$, against which deviations $f_\delta(k,z) := P_{\rm m}(k,z)/P_{\rm m}^{\rm DMO}(k,z)$ are measured, was again halofit (Takahashi et al., 2012) with the SPBK25 parameters, except for a higher $\sigma_8 = 0.81$ to match the increased $\xi_\pm^{(ij)}(\theta)$ amplitude in KiDS-Legacy. We also marginalised over uncertainties in the source redshift distributions, for which we used the shear-independent error model from Wright et al. (2025a) for the calibrated $n_{\rm s}^{(i)}(\chi)$.

For the double power-law approach, we fixed the IA parameters to the best-fitting ΛCDM values. This choice of IA model details does not have a significant impact. Even in the extreme case where we imposed $A_{\rm IA} = 0$, our results typically vary by less than 5%–10%, mostly at high z.

6

6.3 Results and discussion

We applied the regularised deprojection method (SPBK25) to constrain the deviations $f_\delta(k,z)$ in 20 logarithmic bins within $k = 0.01$–$20\,h\,{\rm Mpc}^{-1}$. We used two settings, either averaging over the full redshift range $Z = [0,2]$ or over three separate redshift bins with boundaries $Z_1 = [0, 0.3]$, $Z_2 = [0.3, 0.6]$, and $Z_3 = [0.6, 2]$ to probe for redshift evolution. In the BK24 method, we fitted the double power law to the full angular scale and redshift range of the $\xi_\pm^{(ij)}(\theta)$ data provided.

6.3.1 Matter power spectrum constraints

Figure 6.1 presents the inferred $P_{\rm m}(k,z)$ as a function of scale for the three redshift bins Z_1–Z_3. The Bayesian posterior constraints, $P_{\rm m}(k,z) = f_\delta(k,z)\,P_{\rm m}^{\rm DMO}(k,z)$, are plotted as solid curves for the median with 68th percentile credible intervals (CIs) at the central values, $z_{\rm c}$, of their respective redshift bins: $z_{\rm c} = 0.15$ (purple), 0.45 (blue), and 1.3 (green); the posterior predictive distribution in comparison to the KiDS-Legacy data, and their good match, is shown in Appendix 6.D. As pointed out in SPBK25, most of the lensing signal on the average $f_\delta(k,z)$ originates from lower redshifts within Z_1–Z_3, approximately $z = 0.13$, 0.4, and 0.7, respectively. The dashed curves correspond to the best-fit double power law (with

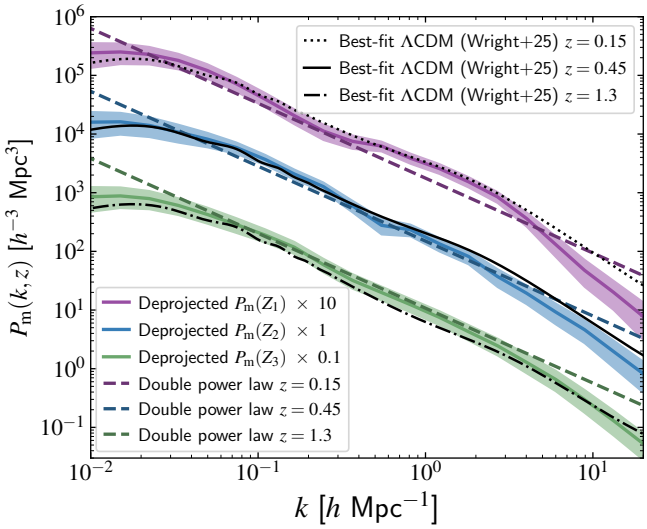


Figure 6.1: Power spectrum constraints in three variants. The solid coloured lines are regularised deprojections with 68% CIs for three redshift bins $Z_1 = [0, 0.3]$, $Z_2 = [0.3, 0.6]$, and $Z_3 = [0.6, 2]$, and the dashed lines are the best fits of a double power law, Eq. (6.2), all interpolated to the centres of Z_1–Z_3. The best-fitting ΛCDM constraints by Wright et al. (2025b) are shown as black curves. For clarity, the curves corresponding to the different redshift bins are scaled by factors of 0.1, 1, or 10. The lensing constraints for Z_3 are mostly from structures near $z \sim 0.7$ (see text).

$k_{\rm piv} = 0.5\ h\ {\rm Mpc}^{-1}$ and $z_{\rm piv} = 0.33$),

$$P_{\rm m}(k,z) = 10^{2.56^{+0.02}_{-0.02}} \left(\frac{1+z}{1+z_{\rm piv}}\right)^{-0.7^{+0.4}_{-0.4}} \left(\frac{k}{k_{\rm piv}}\right)^{-1.28^{+0.02}_{-0.02}} h^{-3}\,{\rm Mpc}^3\,, \tag{6.2}$$

and the additional black curves in the figure correspond to the ΛCDM matter power spectrum obtained with the best-fitting parameters from the KiDS-Legacy cosmic shear inference.

In general, the two reconstruction methods and the ΛCDM fit provide consistent constraints for $P_{\rm m}(k,z)$, albeit within their limitations. For instance, while the regularised deprojection offers maximum flexibility ($0 \leq f_\delta \leq 100$ in 20 k-bins), a strict power-law across the entire range $k = 0.01$–$20\,h\,{\rm Mpc}^{-1}$ is not a good fit to ΛCDM lines at the edges of the k-range, specifically for $k \lesssim 0.03\,h\,{\rm Mpc}^{-1}$ or $k \gtrsim 5\,h\,{\rm Mpc}^{-1}$. However, even the ΛCDM constraint exhibits excessive power in the non-linear regime ($k \gtrsim 3\,h\,{\rm Mpc}^{-1}$) when compared to the posterior CIs of the regularised reconstruction. Regarding redshift evolution, the double power-law indicates a slightly weaker evolution with redshift than the regularised construc-

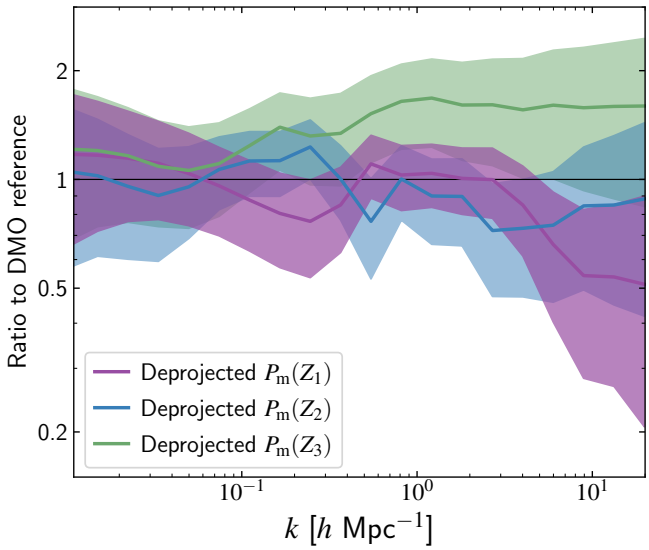


Figure 6.2: Average ratio of the matter power spectrum to the DMO reference within the three redshift bins Z_1–Z_3 derived using the method of regularised deprojection. The shaded regions represent the 68% CIs of the posterior constraints for Z_1 (purple), Z_2 (blue), and Z_3 (green). At large scales, $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$, all three redshift bins align with the expected DMO redshift evolution.

6

tion. However, the double power law is also ill constrained, as shown in Eq. (6.2) and, due to deviations from a simple power law, it also depends on the k-range used in the fit. Nevertheless, the detected amplitude change in the power law improves upon the previous inconclusive KiDS-1000 results (BK24), clearly favouring structure growth with time. A similar conclusion can be drawn from the regularised deprojection, being consistent with the ΛCDM model and its change with z for $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$.

6.3.2 Internal consistency with ΛCDM

In Fig. 6.2, we present the constraints for $f_\delta(k,z)$ for three redshift bins (Z_1: purple, Z_2: blue, Z_3: green). The bins are individually 68% CI consistent with their DMO reference for $k \lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$. At the 1σ level, however, the highest redshift bin, Z_3, still favours slightly more power than the reference spectra at intermediate scales, $k \sim 1\,h\,\mathrm{Mpc}^{-1}$, which might reflect an actual deviation from a DMO scenario on non-linear scales. Our findings are in contrast with the analysis of SPBK25, which identified an internally inconsistent evolution of $P_\mathrm{m}(k,z)$ with the best-fitting ΛCDM cosmology. SPBK25 show that for KiDS-1000, the deviations at medium redshift, Z_2, and high redshift, Z_3, differ significantly from the

KiDS-1000 DMO reference, resulting in a $P_{\rm m}(k,z)$ without significant structure growth in the range $z = 0.4$–0.7, also at large scales. This inconsistency is no longer present in our analysis with KiDS-Legacy.

We attribute the improvement over KiDS-1000 primarily to the enhanced redshift-estimation methodology and the increase in data samples (Wright et al., 2025a). This aligns with the conclusions of Wright et al. (2025b) and Stölzner et al. (2025), who identify the improved redshift distribution estimation as a major contributor to alleviating the S_8 tension. Further factors contributing to consistency with the *Planck* cosmology include the increased survey volume and improved shear calibration. Our analysis also indicates that the refined IA model has a minor beneficial impact on the robustness for the deprojected $P_{\rm m}(k,z)$: it shifts the CI in Fig. 6.2 closer to $f_\delta(k,z) = 1$ by 5%–10% relative to results without IA treatment (not shown). The potential for systematic errors in $f_\delta(k,z)$ stemming from redshift calibration errors and z-dependent IA are further discussed in SPBK25, but these errors contribute less than 10% to the total statistical error.

6.3.3 Small-scale suppression

To boost the signal-to-noise ratio, the 20 grey violin data points in Fig. 6.3 average the constraints for $f_\delta(k,z)$ for a single broad redshift bin, $Z = Z_1 \cup Z_2 \cup Z_3$. The width of the contours represents the posterior probability density. The dark grey horizontal lines show the median; the 68% and 95% CIs are indicated by open boxes and vertical black sticks. At the 1σ level, $f_\delta(k,z)$ is constrained to approximately 20% for each k-bin. The broader CIs indicate that uncertainties increase towards the highest and smallest wavenumbers, where cosmic shear is less constraining. At small scales, $k \gtrsim 3\ h\,\mathrm{Mpc}^{-1}$, there is a clear suppression of the signal. This is better constrained by averaging all data points in the range $k = 3$–$20\,h\,\mathrm{Mpc}^{-1}$. This average, which accounts for the error correlations of $r \lesssim 0.3$ that increase for $k \gtrsim 10\,h\,\mathrm{Mpc}^{-1}$ (Appendix 6.C), yields $\bar{f}_\delta = 0.70 \pm 0.10$. This indicates a 3σ significant detection of suppression of $P_{\rm m}(k,z)$ relative to $P_{\rm m}^{\rm DMO}(k,z)$. We verified this by randomly drawing $\bar{f}_\delta$ values from a null model, $f_\delta(k,z) \equiv 1$, with errors as in our data; 0.3% of these values have $\bar{f}_\delta \leq 0.70$, compatible with a 3σ detection. A similar fit to the results within the individual bins Z_1–Z_3 gives $\bar{f}_\delta = 0.68 \pm 0.18$, 0.71 ± 0.23, and 1.51 ± 0.57 ($|r| \leq 0.4$), respectively. The increasing suppression with time is qualitatively consistent with cosmological hydrodynamical simulations (e.g. Schaller et al., 2025) and tentatively suggests suppression only in Z_1 and Z_2 ($z \lesssim 0.4$), with no suppression in Z_3 ($z \sim 0.7$) at the 1σ level. In addition to statistical uncertainties, we anticipate a conservative systematic uncertainty of $\sim 5\%$ for $\bar{f}_\delta$ because the halofit DMO reference deviates by a factor of 0.95 to 1.05 from the more accurate hmcode2020 (Mead et al., 2021) in the regime of $k = 2$–$10\,h\,\mathrm{Mpc}^{-1}$. Consequently, using hmcode2020 as the DMO refer-

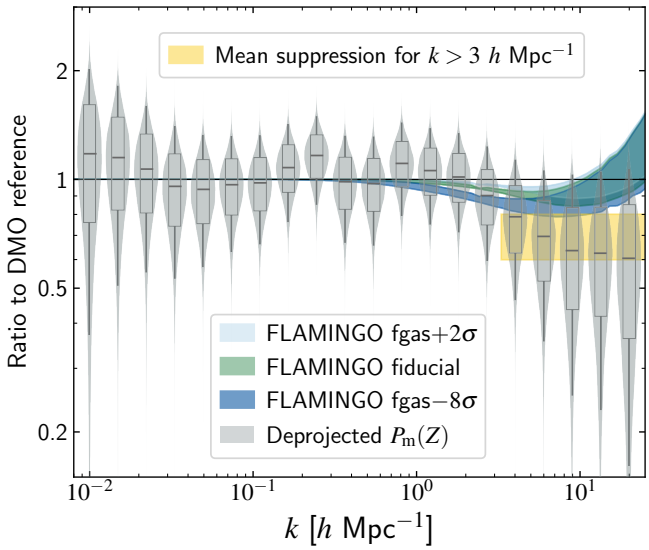


Figure 6.3: Violin plot of the matter power spectrum relative to the DMO spectrum in 20 bands, similar to Fig. 6.2, but for a single broad redshift bin, Z, which combines the smaller bins Z_1–Z_3. The width of the shaded regions represents the posterior probability density; the 68th and 95th percentile CIs about the median are also shown inside the regions (open boxes and sticks, respectively). The green, light blue, and dark blue bands illustrate the suppression predicted by the FLAMINGO cosmological hydrodynamical simulation between $z = 0$ and 1.5 for the fiducial, weak, and strong feedback models, respectively. The yellow band indicates the mean suppression at small angular scales (68% CI), including correlations between the different points (Fig. 6.4), favouring the strong feedback model.

ence changes $\bar{f}_\delta$ by a similar factor.

For theoretical reference, the baryonic suppression signal from the FLAMINGO cosmological hydrodynamical simulations is overplotted in Fig. 6.3 as shaded coloured areas (Schaye et al., 2023; Kugel et al., 2023; Schaller et al., 2025). These simulations are calibrated to (shifts in) the gas fractions of low-redshift groups and clusters. The green band in Fig. 6.3 corresponds to the fiducial FLAMINGO suppression between $z = 0$–1.5. The light and dark blue bands represent variations calibrated to higher ($f_\mathrm{gas} + 2\sigma$) and lower ($f_\mathrm{gas} - 8\sigma$) gas fractions, respectively, which cover the range of predictions by nearly all modern cosmological simulations (see e.g. Fig. 11 in Schaller et al. 2025). The KiDS-Legacy suppression and the FLAMINGO predictions manifest in the same wavenumber regime, indicating that the signal is consistent with the expectation of baryonic feedback. This represents the most significant direct detection of

baryonic feedback from cosmic shear alone (for previous measurements, see Yoon & Jee 2021 and Chen et al. 2023). Therefore, despite the deprojection and systematic uncertainty, $\bar{f}_\delta = 0.70 \pm 0.10\,(\text{stat.}) \pm 0.04\,(\text{sys.})$, there is a preference for the stronger feedback model in KiDS-Legacy. Individually, the weak, fiducial, and strong feedback FLAMINGO variations are respectively 2.6σ, 2.3σ, and 1.9σ discrepant from our inferred value of $\bar{f}_\delta$. A stronger suppression resonates with recent studies that combine kinetic Sunyaev-Zeldovich (SZ) with either WL (Bigwood et al., 2024), clustering of photometric galaxies (Hadzhiyska et al., 2024), galaxy-galaxy lensing (McCarthy et al., 2025), or X-ray (Kovač et al., 2025), and with analyses using thermal SZ and WL (Tröster et al., 2022; Pandey et al., 2025) or fast radio bursts (Reischke & Hagstotz, 2025).

Compared to our regularised deprojection, the KiDS-Legacy results by Wright et al. (2025b) favour weaker feedback, at the level of $f_{\text{gas}} + 2\sigma$ or less (68% CI), for $k \lesssim 3\,h\,\text{Mpc}^{-1}$ (their figures 15 and F.3). This discrepancy is also visible in Fig. 6.1, where the KiDS-Legacy results (black lines using hmcode2020) favour greater power for $k \gtrsim 1\,h\,\text{Mpc}^{-1}$ than our deprojected $P_\text{m}(k,z)$, especially for Z_1 when extrapolated to $10\,h\,\text{Mpc}^{-1}$. We attribute this difference primarily to the employed statistics and their sensitivity to different scales, given the choice of angular cuts. As shown in Wright et al. (2025b, figure 1), our $\xi_\pm^{(ij)}(\theta)$ are sensitive to $\ell \sim 10^4$, whereas their COSEBIs with orders $n \leq 6$ strongly downweight signals beyond $\ell \sim 10^3$. Therefore, compared to the full physics model and the a priori assumptions in Wright et al. (2025b), our agnostic deprojection is sensitive to physics around $k \sim 5\,h\,\text{Mpc}^{-1}$, although it is also more susceptible to systematic errors within this non-linear regime. This interpretation of a sensitivity shift aligns with section 4.3 of Stölzner et al. (2025), who find that bandpowers prefer stronger feedback, peaking near the upper edge of the KiDS-Legacy feedback parameter T_AGN prior. The combination of COSEBIs and bandpowers, which have a window function similar to $\xi_\pm^{(ij)}(\theta)$, peaks at $\log_{10}(T_\text{AGN}/\text{K}) \approx 8.0$. This is, however, dominated by a Bayesian prior that explicitly excludes our suppression level.

6.4 Conclusions

Compared to KiDS-1000, the cosmic shear data from KiDS-Legacy show an improved degree of internal consistency in the context of ΛCDM. In particular, the deprojected growth of structure in the matter power spectrum is 1σ consistent with that of the ΛCDM inference on large scales (Fig. 6.2). The change is primarily attributed to an improved source redshift calibration and the revised image reduction.

The data also reveal deviations from a DMO model on small scales, now probed to higher significance due to the ~ 3.5 times larger survey volume. We measure growth suppression in the matter power spectrum relative to a DMO

model of $1 - \bar{f}_\delta = 30\% \pm 10\%\,(\text{stat.}) \pm 4\%\,(\text{sys.})$ at $k \geq 3\,h\text{Mpc}^{-1}$, with tentative evidence that this suppression is restricted to $z \lesssim 0.4$. The average suppression, detected at 2.8σ significance from cosmic shear data alone, is consistent with baryonic feedback predictions in the FLAMINGO simulations and shows a preference for the stronger feedback variation (Fig. 6.3).

6.5 Data availability

Our constraints for the deprojected $P_\text{m}(k, z)$ and the correlation matrix of its uncertainties are made available at CDS.

References

Amon A., Efstathiou G., 2022, MNRAS, 516, 5355
Asgari M., et al., 2021, A&A, 645, A104
Bigwood L., et al., 2024, MNRAS, 534, 655
Broxterman J. C., Kuijken K., 2024, A&A, 692, A201
Chen A., et al., 2023, MNRAS, 518, 5340
Doux C., Karwal T., 2025, arXiv e-prints, p. arXiv:2506.16434
Fortuna M. C., et al., 2025, A&A, 694, A322
Hadzhiyska B., et al., 2024, arXiv e-prints, p. arXiv:2407.07152
Hildebrandt H., et al., 2017, MNRAS, 465, 1454
Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901
Kovač M., et al., 2025, arXiv e-prints, p. arXiv:2507.07991
Kugel R., et al., 2023, MNRAS, 526, 6103
Kuijken K., et al., 2019, A&A, 625, A2
Lewis A., Challinor A., 2006, Phys. Rep., 429, 1
Li S.-S., et al., 2023, A&A, 670, A100
McCarthy I. G., et al., 2025, MNRAS, 540, 143
Mead A. J., Brieden S., Tröster T., Heymans C., 2021, MNRAS, 502, 1401
Navarro-Gironés D., et al., 2025, MNRAS,
Pandey S., et al., 2025, arXiv e-prints, p. arXiv:2506.07432
Pen U.-L., Lu T., van Waerbeke L., Mellier Y., 2003, MNRAS, 346, 994
Perez Sarmiento K., Laguë A., Madhavacheril M. S., Jain B., Sherwin B., 2025, Phys. Rev. D, 112, 063510
Planck Collaboration et al., 2020, A&A, 641, A6
Preston C., Amon A., Efstathiou G., 2024, MNRAS, 533, 621
Preston C., Rogers K. K., Amon A., Efstathiou G., 2025, MNRAS, 542, 2698
Rauch M., 1998, ARA&A, 36, 267
Reid B. A., et al., 2010, MNRAS, 404, 60
Reischke R., Hagstotz S., 2025, arXiv e-prints, p. arXiv:2507.17742
Reischke R., et al., 2025, A&A, 699, A124
Schaller M., Schaye J., Kugel R., Broxterman J. C., van Daalen M. P., 2025, MNRAS, 539, 1337
Schaye J., et al., 2023, MNRAS, 526, 4978
Seljak U., 1998, ApJ, 506, 64
Simon P., 2012, A&A, 543, A2
Simon P., Porth L., Burger P., Kuijken K., 2025, A&A, 698, A217
Stölzner B., et al., 2025, arXiv e-prints, p. arXiv:2503.19442
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Tröster T., et al., 2022, A&A, 660, A27
Wright A. H., et al., 2024, A&A, 686, A170
Wright A. H., et al., 2025a, arXiv e-prints, p. arXiv:2503.19440
Wright A. H., et al., 2025b, arXiv e-prints, p. arXiv:2503.19441
Ye G., Jiang J.-Q., Silvestri A., 2025, J. Cosmology Astropart. Phys., 2025, 072
Yoon M., Jee M. J., 2021, ApJ, 908, 13

Appendix

6.A Acknowledgements

LP acknowledges support from the DLR grant 50QE2302. HHi is supported by a DFG Heisenberg grant (Hi 1495/5-1), the DFG Collaborative Research Center SFB1491, an European Research Council (ERC) Consolidator Grant (No. 770935), and the DLR project 50QE2305. MA acknowledges support from the UK Science and Technology Facilities Council (STFC) under grant number ST/Y002652/1 and the Royal Society under grant numbers RGSR2222268 and ICAR1231094. MB is supported by the Polish National Science Center through grants no. 2020/38/E/ST9/00395 and 2020/39/B/ST9/03494. CH acknowledges support from the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research, and the UK STFC under grant ST/V000594/1. HHo acknowledges support from the ERC under the European Unions Horizon 2020 research and innovation program with Grant agreement No. 101053992. BJ acknowledges support by the ERC-selected UKRI Frontier Research Grant EP/Y03015X/1 and by STFC Consolidated Grant ST/V000780/1. SSL is receiving funding from the programme "Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of Northrhine Westphalia. SSL acknowledges support from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program with Grant agreement No. 101053992. LM acknowledges the financial contribution from the grant PRIN-MUR 2022 20227RNLY3 The concordance cosmological model: stress-tests with galaxy clusters supported by Next Generation EU and from the grant ASI n. 2024-10-HH.0 Attività scientifiche per la missione Euclid fase E. RR is supported by an ERC Consolidator Grant (No. 770935). MvWK acknowledges the support by UK STFC (grant no. ST/X001075/1), the UK Space Agency (grant no. ST/X001997/1), and InnovateUK (grant no. TS/Y014693/1). Based on data obtained from the ESO Science Archive Facility with DOI: https://doi.org/10.18727/archive/37, and https://doi.eso.org/10.18727/archive/59 and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target;

the University of Padova, and the University Federico II (Naples). 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.

6.B Method details

In this Appendix we provide additional information on the method of regularised deprojection in SPBK25 and that of the double power-law fit in BK24.

6.B.1 Regularised deprojection

For a full account of the regularised reconstruction, we refer to SPBK25, sections 2 and 3. The aim is to have the method invert, without a physical model for the matter power spectrum, the relations

$$\xi_{\pm}^{(ij)}(\theta) = \frac{9H_0^4\Omega_{\rm m}^2}{4c^4} \times \int_0^{\chi_{\rm h}}\int_0^{\infty} \frac{{\rm d}\chi\,{\rm d}\ell\,\ell}{2\pi} \frac{\overline{W}^{(i)}(\chi)\overline{W}^{(j)}(\chi)}{a^2(\chi)} J_{0,4}(\ell\theta)\, P_{\rm m}\left(\frac{\ell+1/2}{f_K(\chi)}, z[\chi]\right) \quad (6.3)$$

with respect to $P_{\rm m}(k,z)$ for a set of tomographic bins $i,j = 1\ldots N_s$. Here the lensing efficiency is defined as

$$\overline{W}^{(i)}(\chi) := \int_{\chi}^{\chi_{\rm h}} {\rm d}\chi'\, n_{\rm s}^{(i)}(z[\chi]) \frac{{\rm d}z(\chi)}{{\rm d}\chi} \frac{f_K(\chi'-\chi)}{f_K(\chi')}\,, \quad (6.4)$$

which varies with each source distribution $n_{\rm s}^{(i)}(z)$, where c is the vacuum speed of light, H_0 is the Hubble parameter, and $\chi_{\rm h}$ is the horizon of the survey. For brevity, we ignore here two additional IA terms, which, under the adopted IA model, only change the projection kernels inside the N_s integrals (6.3) and do not impact the subsequent statements. The projection kernels depend exclusively on $n_{\rm s}^{(i)}(z)$, the comoving angular diameter distance $f_K(\chi)$ (for the curvature scalar K and the comoving distance χ), the matter density $\Omega_{\rm m}$, and the IA parameters. Uncertainties in these four components are propagated into $P_{\rm m}(k,z)$ by varying the projection kernels in repeated reconstructions.

An exact inversion is infeasible due to a low signal-to-noise ratio, confined angular ranges, and the limited number $N_s(N_s+1)/2$ of equations in (6.3). Therefore, as an approximation, we averaged $P_{\rm m}(k,z)$ within $N_z = 1$ or 3 broad z-bins

in radial direction, spanning $z = 0$–2, and within $N_k = 20$ k-bands, ranging between 0.01–$20\,h\,\mathrm{Mpc}^{-1}$. This configuration probes either the full radial average, $N_z = 1$, or the evolution within $N_z = 3$ bins. Outside these k and z ranges, the power spectrum is fixed to a reference power spectrum, $P_\mathrm{m}^\mathrm{DMO}(k,z)$, best fit to the data ($\sigma_8 = 0.81$). All contributions from outside these ranges are encapsulated in the modelled shear signal $\xi_{\pm,\mathrm{fid}}^{(ij)}(\theta)$, which constitutes only a small fraction of the expected total signal $\xi_\pm^{(ij)}(\theta)$. Within these ranges, the amplitudes in the bands can vary freely between a factor of 0 to 100 of $P_\mathrm{m}^\mathrm{DMO}(k,z)$.

However, broad z-bins complicate the interpretation of the averaged power spectrum due to its variation with z. We untangled this by expressing $P_\mathrm{m}(k,z) = f_\delta(k,z)\,P_\mathrm{m}^\mathrm{DMO}(k,z)$ relative to the growing $P_\mathrm{m}^\mathrm{DMO}(k,z)$, and then averaged the more slowly evolving deviations, $f_\delta(k,z)$, instead. Since the adopted DMO reference accounts for dark matter only, $f_\delta(k,z)$ also conveniently averages the deviations of the true $P_\mathrm{m}(k,z)$ from a model without baryon feedback. Mathematically, our statistical model expresses the deviations $f_\delta(k,z)$ across $N_z \times N_k$ bands as the vector $\vec{f}_\delta$. With the linear projections (6.3) presented by the projection matrix $\boldsymbol{X}$, the predicted θ-binned $\xi_\pm^{(ij)}(\theta)$ is given by $\vec{\xi} = \boldsymbol{X}\vec{f}_\delta + \vec{\xi}^\mathrm{fid}$ (assuming fixed projection kernels).

The inversion of the noisy $\vec{\xi} = \boldsymbol{X}\vec{f}_\delta + \vec{\xi}^\mathrm{fid}$ without further constraints leads to strongly oscillating noise for $\vec{f}_\delta$ due to the broad projection kernels and the Bessel functions of first kind, $J_n(x)$, inside the kernels. This issue is mitigated by asserting positive solutions, $f_\delta(k,z) \geq 0$, and by employing a Tikhonov filter (as in SPBK25). This filter attenuates solutions of the band-power $\vec{f}_\delta$ that exhibit strong oscillations in the k-direction, effectively reconstructing a k-smoothed solution. Conversely, oscillations in z-direction for $N_z > 1$ are not filtered to prevent z-smoothing. SPBK25 uses simulated KiDS data for filter settings that strike a balance between degradation by noise and artificial smoothing by the filter.

We inferred the band-power $\vec{f}_\delta$ statistically within a Bayesian framework using the posterior probability density function (PDF) $P(\vec{f}_\delta|\vec{\xi}) \propto \mathcal{L}(\vec{\xi}|\vec{f}_\delta)\,P_\mathrm{hat}(\vec{f}_\delta)\,P_\tau(\vec{f}_\delta)$. Here, the prior PDFs $P_\mathrm{hat}(\vec{f}_\delta)$ and $P_\tau(\vec{f}_\delta)$ implement a uniform prior density, asserting $f_\delta = 0$–100, and the Tikhonov regularisation of $\vec{f}_\delta$. The likelihood function, $\mathcal{L}(\vec{\xi}|\vec{f}_\delta)$, adopts a multivariate Gaussian model for statistical noise in $\vec{\xi}$ with covariance matrix $\boldsymbol{C}$. This covariance incorporates contributions from intrinsic shape noise, cosmic variance, super-sample covariance, and the uncertainty in the multiplicative shear bias. We numerically sampled $P(\vec{f}_\delta|\vec{\xi})$ using a rapidly converging Hamiltonian Monte Carlo algorithm. To propagate statistical errors in $n_\mathrm{s}^{(i)}(z)$, the angular diameter distance and Ω_m, and IA correlations, we combined different Monte Carlo chains where the projection kernels are randomly drawn from a model of kernel uncertainties. The inferred $P_\mathrm{m}(k,\bar{z}) = \bar{f}_\delta(k)\,P_\mathrm{m}^\mathrm{DMO}(k,\bar{z})$ at the centres, $\bar{z}$, of three redshift bins for the averaged $\bar{f}_\delta(k)$ is shown in Fig. 6.1 with statistical uncertainties.

6.B.2 Double power law

For the shape of the matter power spectrum, the analysis of BK24 assumes a double power law of the form

$$P_{\rm m}(k,a) = A\left(\frac{k}{k_{\rm piv}}\right)^{p}\left(\frac{a}{a_{\rm piv}}\right)^{m}, \tag{6.5}$$

where the pivot points $k_{\rm piv}$ and $a_{\rm piv}$ are determined by minimising the covariance between the different model parameters. The free parameters $\log_{10}(A/h^{-3}\,{\rm Mpc}^3)$, p, and m are inferred through a Bayesian inference by assuming the flat priors $\log_{10}(A/h^{-3}\,{\rm Mpc}^3) \in [0,10], p \in [-2,0.5]$, and $m \in [-5,5]$.

Additional freedom in modelling is chosen to replicate the choices made in the Stage III cosmic shear inferences. As such, the free parameters of the IA model, except for $A_{\rm IA}$, the integration limits of the different two-point statistics, and the scale cuts are taken from the fiducial Stage III cosmic shear inferences. The background cosmology, used to compute the linear growth factor and redshift-comoving distance relation, however, is chosen to be the best-fitting ΛCDM cosmology from Planck Collaboration et al. (2020). Changing this to the best-fitting KiDS-1000 cosmology from Asgari et al. (2021) has a negligible impact on the results.

As shown in Fig. 6.1, the maximum-likelihood solutions (dashed lines) are a reasonable fit to the regularised reconstruction within $0.03 \lesssim k \lesssim 5\,h\,{\rm Mpc}^{-1}$. The systematically lower amplitude of the fit in this k-regime, compared to the regularised construction, is explained by increasing deviations from self-similar evolution outside this regime.

6.C Correlation matrix of deprojection uncertainties for a single redshift bin

The deprojection of the tomographic $\xi_{\pm}^{(ij)}(\theta)$ into $f_\delta(k,z)$, with measurement noise localised in θ, produces correlated uncertainties across all k and z bins in $f_\delta(k,z)$ and, consequently, in $P_{\rm m}(k,z)$ (SPBK25). Figure 6.4 displays the correlation matrix for the results in Fig. 6.3 with single z-bin Z. The Pearson correlation coefficients, r, of the deprojected $P_{\rm m}(k,z)$ between 20 logarithmically spaced bins for $k = 0.01$–$20\ h\ {\rm Mpc}^{-1}$ are presented (from bottom left to top right). The blue colours indicate that estimates at different scales are at most moderately correlated, typically with $|r| \lesssim 0.3$. The smallest and largest scales exhibit the strongest correlation with their directly adjacent bins. The correlation matrix for the deprojection with three separate z-bins (Z_1 to Z_3) is similar to that presented in figure 5 of SPBK25.

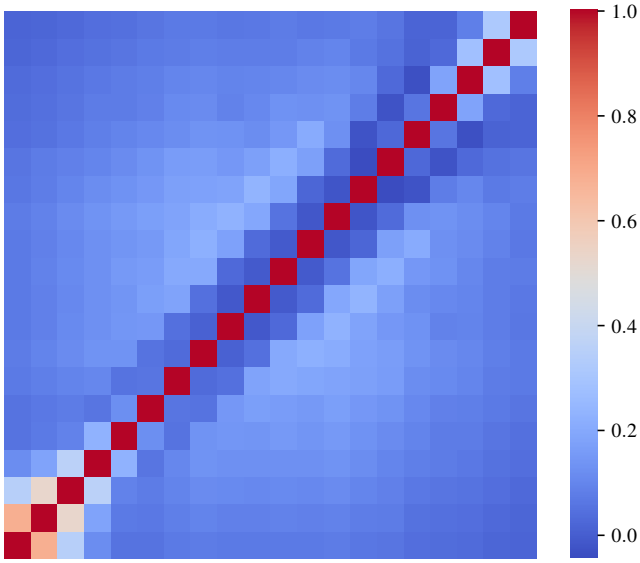


Figure 6.4: Matrix of Pearson correlation coefficients for the deprojection with a single z-bin, Z, for 20 logarithmic k-bins in the range $k = 0.01 - 20\ h\ \mathrm{Mpc}^{-1}$. The scale k increases from bottom left to top right.

6.D Posterior predictive distribution

Figure 6.5 is the posterior predictive distribution (PPD) of $P_\mathrm{m}(k,z)$ deprojected into three separate redshift bins Z_1–Z_3 (solid coloured curves in Fig. 6.1). The 68% and 95% CIs of the PPD (in blue) assume a fixed lensing kernel and no IA uncertainties, which, if accounted for, modestly increase the CIs by approximately 5%. The solid red line represents the prediction from the DMO reference $P_\mathrm{m}^\mathrm{DMO}(k,z)$, adjusted in the Bayesian deprojection by a factor $f_\delta(k,z)$, presented in Fig. 6.2, to match the observed data (correlated black data points with 1σ uncertainty). Although an additional sixth source redshift bin is included, compared to the previous KiDS-1000 analysis in SPBK25, our PPD consistently provides a good description of all $\xi_\pm^{(ij)}(\theta)$ data points, although a few notable outliers are present, such as for $\xi_\pm^{(33)}(\theta)$, $\xi_+^{(13)}(\theta)$, $\xi_-^{(12)}(\theta)$, or $\xi_+^{(26)}(\theta)$.

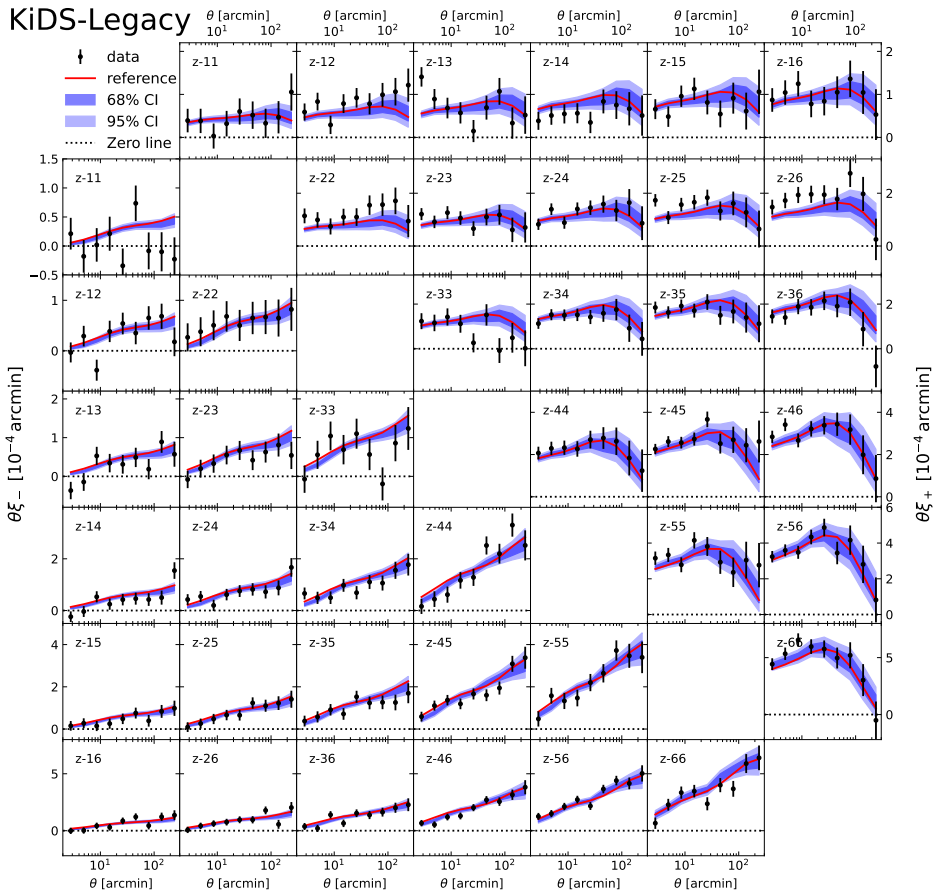


Figure 6.5: Posterior predictive distribution (68% and 95% CIs as blue regions) of the deprojected $P_{\rm m}(k,z)$ in Fig. 6.1 relative to the black data points – $\theta\,\xi_-^{(ij)}(\theta)$ in the lower left panels and $\theta\,\xi_+^{(ij)}(\theta)$ in the upper right panels – for a combination (ij) of tomographic source bins, denoted as '$z-ij$' inside the panels. The red lines represent the reference model, $P_{\rm m}^{\rm DMO}(k,z)$, employed in the regularised deprojection.

Nederlandse samenvatting

Introductie

Observationele kosmologen proberen de inhoud en evolutie van het universum te begrijpen en te beschrijven aan de hand van metingen. Voorbeelden van observaties die hiervoor gebruikt kunnen worden zijn de kosmische microgolfachtergrondstraling (CMB), fotonen die sinds een paar honderdduizend jaar na de oerknal vrij door het universum voortbewegen, toen het hele universum nog een extreem warm geheel was, baryonische akoestische oscillaties, een afdruk in de verdeling van sterrenstelsels afkomstig uit dezelfde tijd als de CMB, en supernovae, waarvan de intrinsieke lichtkracht een manier biedt om absolute afstanden te kalibreren.

Nog een andere methode, die bijzonder krachtig is voor het bestuderen van de hoeveelheid en verdeling van alle materie, namelijk de combinatie van donkere en gewone materie, is zwakke zwaartekrachtlenswerking. Dit effect is een gevolg van het feit dat materie de ruimte kromt, zoals beschreven door Einstein in zijn algemene relativiteitstheorie. Als consequentie hiervan worden lichtdeeltjes een klein beetje afgebogen door variaties in de verdeling van materie in het kosmische web, de grootste niet-homogene structuur in het heelal. De afbuiging van deze fotonen uit zich in kleine vervormingen in de afbeeldingen van verre sterrenstelsels. Door een groot aantal sterrenstelsels statistisch te analyseren, kunnen we een idee krijgen van de hoeveelheid materie en van (de evolutie van) de materieverdeling. Zwakke zwaartekrachtlenswerking is een van de hoofdmetingen die ESA's *Euclid*-satelliet gebruikt om de wetten van het universum te bestuderen.

Om de verdeling van materie correct te beschrijven, is het belangrijk om alle fysische processen die deze verdeling beïnvloeden goed te begrijpen. Voor het kosmische web is de belangrijkste kracht de zwaartekracht, die langzaam de structuur van het universum vormt. Er zijn echter ook andere fysische processen die we niet kunnen verwaarlozen met de precisie van de huidige instrumenten, laat staan met de verwachte resultaten van de *Euclid*-satelliet.

Voor studies die gebruikmaken van zwakke zwaartekrachtlenswerking zijn baryonische processen een van de processen met de grootste onzekerheid, waarvan de belangrijkste de feedback van superzware zwarte gaten en de explosies van supernovae zijn. Deze processen verhitten en verplaatsen gas tot buiten de

sterrenstelsels waarin ze plaatsvinden. Op de schalen relevant voor *Euclid* zorgt deze activiteit voor minder klontering van materie dan je zou verwachten op basis van zwaartekracht alleen.

Om met zwakke zwaartekrachtlenswerking de juiste evolutiemodellen van het universum te bepalen, is het zaak om deze baryonische fysica dus ook goed te begrijpen. Hoewel de baryonische effecten erg interessant kunnen zijn omdat ze ons kunnen vertellen over de eigenschappen van bijvoorbeeld superzware zwarte gaten, hun jets en de fysica die hiermee gepaard gaat, worden ze door de huidige zwakke zwaartekrachtlensstudies voorlopig vooral gezien als een storende factor. De meest zelfconsistente manier om deze processen te bestuderen is door middel van kosmologische hydrodynamische simulaties. Dit zijn simulaties die proberen het universum zo nauwkeurig mogelijk na te bootsen door de belangrijkste fysische processen te modelleren. In deze simulaties kan vervolgens de sterkte van onzekere fysische processen worden aangepast om een idee te krijgen van niet alleen de absolute impact, maar ook van de bijbehorende onzekerheid.

Dit proefschrift

Hoofdstuk 2: De invloed van baryonische fysica op het aantal zwakke zwaartekrachtlenswerkingpieken

In hoofdstuk 2 gebruiken we de FLAMINGO-simulaties om de invloed van baryonische processen op een specifieke zwakke zwaartekrachtlenswerkingstatistiek te bestuderen. Waar eerdere simulaties vaak slechts één voorspelling maken voor de impact van deze fysische effecten, varieert FLAMINGO de sterkte van deze processen op een systematische manier. Hierdoor kan de invloed van het veranderen van de onderliggende kosmologie ongeveer gescheiden worden van de impact van de superzware zwarte gaten. In dit hoofdstuk kijken we naar de invloed van deze baryonische processen op de zwakke zwaartekrachtlenswerkingpieken, de lokale maxima in het zwakke zwaartekrachtlenswerkingconvergentieveld. Met behulp van de FLAMINGO-lichtkegels modelleren we het verwachte signaal van *Euclid* in de verschillende simulaties. De verschillen tussen de simulaties die met verschillende kosmologische parameters zijn uitgevoerd, zijn groter dan die met verschillende feedback. Er zit dus interessante kosmologische informatie in de hoogteverdeling van zwakke zwaartekrachtlenswerkingpieken. We zien echter ook dat de impact van baryonische fysica niet genegeerd kan worden bij de analyse van *Euclid*, en dit zal dus nauwkeurig meegenomen moeten worden in toekomstige analyses met echte observaties.

Hoofdstuk 3: Kosmologie met de roodverschuivingsverdeling van zwakke zwaartekrachtlenswerkingpieken

Hoofdstuk 3 is een vervolgstudie op de resultaten van hoofdstuk 2. In dit hoofdstuk kijken we naar dezelfde objecten, zwakke zwaartekrachtlenswerkingpieken, maar nu niet naar hun hoogteverdeling, maar naar hun roodverschuivingsverdeling, die kwantificeert hoe ver de objecten die de pieken genereren van de waarnemer afzitten. We gebruiken wederom de FLAMINGO-simulaties om iedere hoge piek te proberen te paren met een individuele massieve halo in de simulatie. Hiermee laten we zien dat het merendeel van de hoge pieken die *Euclid* zal meten overeenkomt met een enkele halo langs de gezichtslijn. Daarnaast laten we zien dat de vorm van de roodverschuivingsverdeling wel afhangt van kosmologie, maar in een stuk mindere mate van baryonische feedback. Hierdoor is het een geschikte meting om kosmologie mee te analyseren. Aangezien de piekhoogteverdeling, die in hoofdstuk 2 bestudeerd werd, sterker van baryonische feedback afhangt, kunnen de twee verdelingen worden gecombineerd om de sterkte van de baryonische feedback af te schatten, terwijl ook de kosmologische parameters worden bepaald, ook wel zelfkalibratie genoemd.

Hoofdstuk 4: Het kosmologisch onderscheidend vermogen van de roodverschuivingen, hoogtes en hoekclustering van zwakke zwaartekrachtlenswerkingpieken

Hoofdstuk 4 is een vervolg op hoofdstukken 2 en 3. In dit hoofdstuk kijken we wederom naar de zwaartekrachtlenswerkingpieken, maar nu naar hun hoogteverdeling, roodverschuivingsverdeling en hoekclustering. Hoekclustering kwantificeert de relatieve verdeling van de pieken ten opzichte van elkaar aan de hemel. We gebruiken de FLAMINGO-hyperkubus, een set simulaties waarin 10 kosmologische parameters tegelijk gevarieerd worden, en we modelleren wederom het *Euclid*-signaal in deze simulaties. We bestuderen het vermogen van de verschillende piekeigenschappen om de kosmologische parameters te bepalen. We vinden dat de roodverschuivingsverdeling van de pieken over het algemeen individueel de meest precieze resultaten oplevert. Wanneer we de verschillende piekverdelingen combineren, zien we echter dat het gecombineerde resultaat beter is dan je zou verwachten op basis van de individuele resultaten alleen. De vergelijking toont nogmaals aan dat de zwakke zwaartekrachtlenswerkingpieken een waardevolle statistiek kunnen zijn om de *Euclid*-data mee te analyseren, analoog aan traditionele clusternummerstudies.

Hoofdstuk 5: Eerste stap naar de reconstructie van het materievermogensspectrum met fase III zwakke zwaartekrachtlenswerking waarnemingen

Een veelbelovende methode voor het analyseren van verschillende kosmologische observabelen is de reconstructie van het materievermogensspectrum, dat de verdeling van materie over verschillende schalen beschrijft. Een reconstructie van deze verdeling kan inzicht geven in op welke kosmische tijden en op welke kosmologische schalen metingen verschillen, maar kan ook helpen bij het uitsluiten van specifieke modellen en als een consistentietest dienen voor kosmologische studies. In dit hoofdstuk gebruiken we de openbaar beschikbare data van de drie fase III zwakke zwaartekrachtlenswerkingteams. We nemen een simpel model aan voor het materievermogensspectrum, een dubbele machtswet in schaalfactor en golfgetal. Hoewel het model slechts empirisch is en de onzekerheden in de metingen groot zijn, vinden we dat de evolutie met kosmische tijd inconsistent is met de Λ koude donkere materie (KDM) studies met dezelfde data.

Hoofdstuk 6: Reconstructie van het materievermogensspectrum met KiDS-Legacy: verbeterde interne consistentie en voorkeur voor sterke baryonische feedback

In hoofdstuk 6 herhalen we de analyse uit hoofdstuk 5, maar nu met de nieuwste data van het KiDS-team. Daarnaast maken we gebruik van een geavanceerdere regularisatiemethode die afwijkingen meet ten opzichte van een universum met alleen donkere materie. We vinden dat het best passende model voor het materievermogensspectrum nu een stuk dichter bij de verwachting van het ΛKDM-model ligt. Daarnaast zien we op kleine schaal een onderdrukking van het signaal die overeenkomt met de verwachting van sterke baryonische processen.

Publications

First author

1. *The FLAMINGO project: Baryonic impact on weak gravitational lensing convergence peak counts*
   Jeger C. Broxterman, Matthieu Schaller, Joop Schaye, Henk Hoekstra, Konrad Kuijken, John C. Helly, Roi Kugel, Joey Braspenning, Willem Elbers, Carlos S. Frenk, Juliana Kwan, Ian G. McCarthy, Jaime Salcido, Marcel P. van Daalen, and Bert Vandenbroucke
   Published in MNRAS, 2024, 529, 2309

2. *First step toward matter power spectrum reconstruction with Stage III weak gravitational lensing surveys*
   Jeger C. Broxterman and Konrad Kuijken
   Published in A&A, 2024, 692, A201

3. *The FLAMINGO project: cosmology with the redshift dependence of weak gravitational lensing peaks*
   Jeger C. Broxterman, Matthieu Schaller, Henk Hoekstra, Joop Schaye, Robert J. McGibbon, Victor J. Forouhar Moreno, Roi Kugel, and Willem Elbers
   Published in MNRAS, 2025, 538, 755

4. *Matter power spectrum reconstruction with KiDS-Legacy: Improved internal ΛCDM consistency and preference for strong baryonic feedback*
   Jeger C. Broxterman, Patrick Simon, Lucas Porth, Konrad Kuijken, Angus H. Wright, Marika Asgari, Maciej Bilicki, Catherine Heymans, Hendrik Hildebrandt, Henk Hoekstra, Benjamin Joachimi, Shun-Sheng Li, Matteo Maturi, Lauro Moscardini, Mario Radovich, Robert Reischke, and Maximilian von Wietersheim-Kramsta
   Published in A&A, 2025, 703, L3

   VizieR Online Data Catalog: Matter power spectrum reconstruction (Broxterman+, 2025)
   VizieR Online Data Catalog 370, J/A+A/703/L3

5. *Cosmological constraining power of the redshifts, heights, and angular clustering of weak gravitational lensing peaks*
   Jeger C. Broxterman, Matthieu Schaller, Ian G. McCarthy, Willem Elbers, John Helly, Henk Hoekstra, Konrad Kuijken, Jaime Salcido, Joop Schaye, Naomi Schutte, and Elena Sellentin
   Published in MNRAS, 2026, 553, stag1905

Co-author

1. *Formation of phenylacetylene and benzocyclobutadiene in the ortho-benzyne + acetylene reaction*
   Morgan N. McCabe, Patrick Hemberger, Dario Campisi, **Jeger C. Broxterman**, Engelbert Reusch, Andras Bodi, and Jordy Bouwman
   Published in PCCP, 2022, 24, 1869

2. *Non-linear CMB lensing with neutrinos and baryons: FLAMINGO simulations versus fast approximations*
   Amol Upadhye, Juliana Kwan, Ian G. McCarthy, Jaime Salcido, John C. Helly, Roi Kugel, Matthieu Schaller, Joop Schaye, Joey Braspenning, Willem Elbers, Carlos S. Frenk, Marcel P. van Daalen, Bert Vandenbroucke, and **Jeger C. Broxterman**
   Published in MNRAS, 2024, 529, 1862

3. *The FLAMINGO project: baryon effects on the matter power spectrum*
   Matthieu Schaller, Joop Schaye, Roi Kugel, **Jeger C. Broxterman**, and Marcel P. van Daalen
   Published in MNRAS, 2025, 539, 1337

4. *Euclid: An emulator for baryonic effects on the matter bispectrum*
   Pierre A. Burger, Giovanni Aricò, Laila Linke, Raul E. Angulo, **Jeger C. Broxterman** *et al.*
   Published in A&A, 2025, 705, A170

5. *The resummation model in FLAMINGO: precisely predicting matter power suppression from observed halo baryon fractions*
   Marcel P. van Daalen, Ioannis Koutalios, **Jeger C. Broxterman**, Bart J.H. Wolfs, John C. Helly, Matthieu Schaller, and Joop Schaye
   Published in MNRAS, 2026, 545, staf2086

6. *Euclid preparation LXXXV. Towards a DR1 application of higher-order weak lensing statistics*
   Simone Vinciguerra, Filippo Bouchè, Nicolas Martinet *et al.*
   Published in A&A, 2026, 707, A235

7. *FLAMINGO: Baryonic effects on the weak lensing scattering transform*
   Mariia Marinichenko, Marcel P. van Daalen, Elena Sellentin, **Jeger C. Broxterman**, Matthieu Schaller, and Joop Schaye
   Published in MNRAS, 2026, 546, stag190

8. *The FLAMINGO simulations data release*
   John C. Helly, Robert J. McGibbon, Joop Schaye, Matthieu Schaller, William McDonald, Joey Braspenning, **Jeger C. Broxterman**, Emily E. Costello, Willem Elbers, Carlos S. Frenk, Adrian Jenkins, Roi Kugel, Ian G. McCarthy, Jaime Salcido, Marcel P. van Daalen, Bert Vandenbroucke, Tianyi Yang
   Published in Astron. Comput., 2026, 57, 101159

9. *Ray-traced weak lensing convergence in screened modified gravity theories*
   Mattia Pantiri, Matthieu Schaller, Alessandra Silvestri, **Jeger C. Broxterman**, and Joop Schaye
   Published in MNRAS, 2026, 551, stag1523

10. *cloelib: A Flexible Python Library for Computing Cosmological Observables in the Euclid Era*
    Marco Bonici, Guadalupe Cañas-Herrera, Pedro Carrilho, Santiago Casas, Chiara Moretti, Andrea Pezzotta *et al.*
    Submitted to JOSS, arXiv:2605.23839

Curriculum Vitae

I was born on Wednesday, March 24th, 1999, in the hospital in Nieuwegein, the Netherlands, together with my twin sister. I grew up in Geldermalsen, in the same house where my mother had grown up some 30 years earlier. At the age of 4, we went to the Oranje Nassau School in the same town. At this age, I aspired to be a professional runner, likely because my height gave me an unfair advantage over my classmates. As Geldermalsen only had a secondary school offering pre-vocational secondary education, I had to choose a school in one of the neighbouring cities. After some consideration, I chose to attend the same secondary school as my sisters, the Koningin Wilhelmina College in Culemborg, a 10-kilometre bike ride from our parental home. When we had to choose our secondary school specialisations, it was clear I had an affinity for the exact sciences. Therefore, I chose the science specialisations, and from the fourth year onwards, I attended U-Talent, a pre-university programme in which secondary school students visit Utrecht University for two days per month to attend additional science courses and gain initial insight into academic research. Here, for my *profielwerkstuk*, I got the opportunity to study the reproductive strategy of vegetation in tidal marshes, including a field trip to Yerseke for a couple of days to dig up common cordgrass under the watchful eye of the Borssele nuclear power plant.

When the time came to choose what I wanted to do after secondary school, I visited several Dutch universities, quickly realising I did not want to attend a technical university. Despite my interest in a range of topics, I decided to study physics in Leiden because of my continued interest in the exact sciences, and because it seemed like a well-established field with excellent career prospects. I decided to pursue a joint bachelor's degree in astronomy because it was an easy option that seemed to require little additional work, and because high school had always gone well for me. During these years, I slowly realised I enjoyed the astronomy courses much more, the highlight of which was, of course, the observing trip to La Palma, and therefore I decided to enrol in the Astronomy & Cosmology master specialisation in Leiden. Alongside my studies, I have also pursued other interests by taking additional university courses in art history, econometrics, and, later, Spanish. During my master's degree, I became particularly fascinated by our understanding of the Universe and by how well a relatively simple theory can predict so many observational probes. I also greatly enjoyed the research

projects I undertook, particularly my bachelor's research project on the formation of aromatic molecules under interstellar conditions in Jordy Bouwman's lab, and my master's research project on the modelling of cosmic shear data. The latter allowed me to witness the day-to-day work of analysing data from a state-of-the-art cosmological survey, as I attended a KiDS busy week at *Kasteel Oud-Poelgeest* in Oegstgeest.

At the end of my master's programme, I was still amazed by observational cosmology and decided to apply for a PhD position at Leiden. As you may imagine, I was thrilled when I got the opportunity to contribute actively to this field by carrying out a PhD at the Lorentz Institute for Theoretical Physics and Leiden Observatory, in the galaxy formation and simulation and weak gravitational lensing groups, working together with the observational cosmologists from KiDS and *Euclid* and the simulators from FLAMINGO. During my PhD, I have focused on modelling *Euclid*-like cosmic shear signals in the FLAMINGO simulations, with particular emphasis on the properties of a single higher-order observable, namely weak-lensing peaks. Additionally, I explored a promising method for reconstructing the matter power spectrum from weak gravitational lensing data.

I have also had the opportunity to be directly involved in the exploitation of ESA's *Euclid* mission. In particular, I was part of the higher-order weak-lensing statistics work package of the weak-lensing science working group and the *Euclid* Consortium (EC) news team of the EC internal communications group. I was also part of the local organising committee for the EC annual meeting in Leiden in March 2025, which was attended by ≈ 500 people, who collectively sang 'Happy Birthday' to me.

Given that I enjoy it and because I believe it is important to communicate our findings, methods, and considerations to the general public, who ultimately fund our endeavours, I have been involved in several public outreach projects. With Astronomy on Tap Leiden, we organised monthly events with two layman's talks in a pub in Leiden. As part of the *Leven in het Heelal* programme, each semester we visited a primary school for a 5-week programme in which children designed their own alien to learn about astronomy and sustainability. With Plants & Planets, we contributed to a new exhibition at the Old Observatory, aiming to bring together botanists from the Hortus and astronomers from Leiden Observatory. Finally, I have also given outreach talks at Astronomy on Tap and to an amateur astronomy enthusiast group from Zwolle.

After careful consideration, I have decided not to pursue a further career in academia. Among other reasons, diminishing enthusiasm for current cosmological inferences and a desire to try out and learn something new led me to conclude that now is the right time for me to change gears. Nevertheless, I remain very grateful for the opportunities I have been given, and I have enjoyed them every step of the way.

Acknowledgements

Whenever I start reading a new book, I always first read the final line. Although rarely revealing any pivotal details, it somehow feels like an illicit glimpse into the magic of the story yet to come. Writing this section feels like having the opportunity to write the final line of my astronomical story. In autumn 2017, I arrived in Leiden as an ignorant boy who wanted to study physics, and who only studied astronomy because it was an easy joint option. Luckily, I was slowly captivated by astronomy, particularly cosmology, and by the incredible detail with which we understand the Universe.

I have always felt very privileged by the opportunities I have had, and am therefore incredibly thankful to have had the chance to eventually arrive at this thesis. Let alone for simply having been born in a time and place that has allowed me to spend so much of my time studying and contributing to astronomy, while not having to bother with any of the basic human needs, or simply for the sheer luck that I have not had to worry about my health and my brain is able to comprehend maths and physics better than most. For this, I cannot feel anything but the deepest feelings of gratitude. A great feeling of thankfulness also goes out to all those who have worked to keep the physics and astronomy institutes running as smoothly as they have and to make it such a pleasant place to be for all these years.

In this short section, I can never do justice to all who have advised, supported, encouraged, distracted, or otherwise helped me. Nevertheless, if you have made it here, I hope you know I have thought of you while writing this section, and you are very much appreciated.

I don't think there was a better place in the world to pursue my PhD research topic over the past years than Leiden. The unique combination of the expertise on weak lensing and cosmological hydrodynamical simulations, with the FLAMINGO simulations having been run the year before I started, and *Euclid*'s launch at the end of my first year, could not have been better. Matthieu, thank you for all our conversations, your guidance, your seemingly endless computer knowledge, your encouragement and support, and for always taking the time to help if possible. Joop, thank you for leading by true scientific example. Your knowledge, attitude, and leadership have been nothing but pristine. Henk, thank you for your relentless optimism, drive, and enthusiasm, and for being

an advocate for scientific outreach. Koen, thank you for introducing me to observational cosmology in the first place and for your calm yet thorough manner. Without my master's research project with you, I probably would not have been here. Additional gratitude goes out to Patrick Simon, with whom I collaborated closely on the results presented in Chapter 6, a timely project I greatly enjoyed.

To the many incredible scientists who have passed through the simulation group – Bert, Elia, Evgenii, Filip, Joey, Lýdia, Maarten, Orestis, Rob, Roi, Victor, Will, Yanling, Zorry – and the lensing group – Aniruddh, Casper, Christos, David, Dennis, Divya, Elisa, Gimena, Guada, Jinyi, Juan, Lammim, Maria, Marloes, Mijin, Petra, Qing, Shun-Sheng, Thomas, Yun-Hao, Zeynab – it has been an incredible privilege to hear you talk about your research and discuss science with you. I hope you get where you want to be.

Additionally, I want to thank all who have been part of the Astronomy on Tap, Life in the Universe project, and Plants & Planets. I think our work has been very valuable and hope to see you continue it for many more years.

On a personal level, there remain more people to thank than the word limit allows me. I can only reconfirm my appreciation for having been part of each other's lives. Whether all the way back from school, university, student life, as flatmates, teammates, colleagues, or simply as friends, thank you. In particular, Niels and Thom, thank you for standing by my side today as my paranymphs, and for having done so for many years, and Henk and Stan, and your love for cooking, who had to put up with me as a flatmate whilst working on this thesis.

Finally, to my family, in particular my mother and father, thank you for your continued support and interest, for providing me with solid ground from which to explore my interests, for the trust that I will figure out which path is mine, and for always giving me the space and confidence to do so.

That leaves me to conclude with the motto on my great-grandmother Goes's plate, which stands on a shelf in my parents' home and hopefully sums up the next stage of my life: *Na druk, geluk.*