

# **Extracting cosmological information from small scales in weak gravitational lensing data**

**José Manuel Zorrilla Matilla**

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
under the Executive Committee  
of the Graduate School of Arts and Sciences

**COLUMBIA UNIVERSITY**

2020

© 2020

José Manuel Zorrilla Matilla

All Rights Reserved

## **Abstract**

# **Extracting cosmological information from small scales in weak gravitational lensing data**

José Manuel Zorrilla Matilla

This work is concerned with how to extract information encoded in small scales of non-Gaussian fields, with the purpose of learning about cosmology using weak gravitational lensing. We do so by comparing different methods on simulated data sets. The topic is relevant, for upcoming galaxy surveys will map the late evolution of the matter density field, which is non-Gaussian, with an unprecedented level of detail, and any improvement on the analysis techniques will increase the experiments' scientific return.

First, we investigate some non-Gaussian observables used in the weak lensing community. We analyze to what extent they are sensitive to the background expansion of the universe, and to what extent to the evolution of the structures responsible for the lensing. We then focus our attention on one such statistic, lensing peaks, and assess the performance of a simple halo-based model that has been proposed to forecast their abundance. We find some shortcomings of that semi-analytic approach, and proceed to review some minimal requirements for numerical simulations used to forecast non-Gaussian statistics, to reduce their computational cost while fulfilling the accuracy and precision required by future experiments.

Second, we propose a novel measurement, that of the temperature dipole induced on the cosmic microwave background induced by the rotation of ionized gas around galaxies, as an additional observation to help constrain the distribution of baryonic matter on the smallest scales probed by WL experiments. The uncertainty in this distribution is a major theoretical systematic for future surveys.

Third, we show how deep neural networks can be used to map pixel-level data into the cosmological parameters of interest, by-passing the previous compression step of measuring pre-designed statistics. We provide the first (simulation-based) credible contours based on neural networks applied to weak lensing data, and discuss how to interpret these models.

## Table of Contents

List of Tables . . . . .	vi
List of Figures . . . . .	x
Acknowledgments . . . . .	xxii
Dedication . . . . .	xxiii
Chapter 1: Introduction . . . . .	1
1.1 Weak lensing cosmology . . . . .	1
1.1.1 Weak lensing fundamentals . . . . .	2
1.1.2 Experimental results . . . . .	4
1.2 Cosmological inference based on WL measurements . . . . .	5
1.2.1 WL statistics . . . . .	7
1.3 Deep learning applied to weak lensing cosmology . . . . .	9
1.3.1 Deep learning basics . . . . .	10
1.3.2 Applications of deep learning to WL . . . . .	17
1.4 Structure of dissertation . . . . .	18
Chapter 2: Geometry and growth contributions to cosmic shear observables . . . . .	21
2.1 Introduction . . . . .	21



2.2	Disentangling geometry from growth in simulations . . . . .	23
2.2.1	Simulating weak lensing maps . . . . .	23
2.2.2	Isolating the effect of geometry vs. growth . . . . .	26
2.3	Sensitivity to $\Omega_m$ and $w$ . . . . .	27
2.3.1	Power spectrum . . . . .	28
2.3.2	Equilateral bispectrum . . . . .	30
2.3.3	Lensing peaks . . . . .	32
2.3.4	Minkowski functionals . . . . .	34
2.4	Impact on parameter inference . . . . .	36
2.5	Discussion . . . . .	41
2.6	Conclusions . . . . .	44
Chapter 3: Do dark matter halos explain lensing peaks? . . . . .		48
3.1	Introduction . . . . .	48
3.2	Predicting peak counts . . . . .	50
3.2.1	N-body simulations . . . . .	51
3.2.2	Camelus . . . . .	52
3.2.3	Parameter inference . . . . .	52
3.3	Results . . . . .	57
3.4	Discussion . . . . .	66
3.5	Conclusions . . . . .	75
Chapter 4: Optimizing simulation parameters for weak lensing analyses involving non-Gaussian observables . . . . .		77

4.1	Introduction . . . . .	77
4.2	Methods . . . . .	78
4.2.1	Simulating convergence maps . . . . .	78
4.2.2	Assessing the impact of hyper-parameters . . . . .	81
4.2.3	Hyper-parameter configurations . . . . .	83
4.2.4	Observables . . . . .	84
4.3	Results and discussion . . . . .	89
4.3.1	Lens plane thickness . . . . .	89
4.3.2	Mass resolution . . . . .	96
4.4	Conclusions . . . . .	101
Chapter 5: Probing gaseous galactic halos through the rotational kSZ effect . . . . .		105
5.1	Introduction . . . . .	105
5.2	Modeling the rotational kSZ (rkSZ) signal from galaxies . . . . .	106
5.2.1	The rkSZ imprint on the CMB . . . . .	106
5.2.2	Galactic atmospheres: electron density . . . . .	107
5.2.3	Galactic atmosphere: kinematics . . . . .	110
5.3	Characterizing the observed rkSZ signal . . . . .	112
5.3.1	Aperture filter . . . . .	114
5.3.2	Matched filter . . . . .	115
5.4	Measurement signal-to-noise and required number of galaxies . . . . .	115
5.5	Stacking <i>Planck</i> data at the positions of MaNGA galaxies . . . . .	122
5.5.1	Galaxy data: MaNGA . . . . .	122

5.5.2	CMB data: <i>Planck</i>	124
5.5.3	Stacking	124
5.6	Discussion	129
5.6.1	Model and observational uncertainties	129
5.6.2	Measurement uncertainties	132
5.6.3	Detection feasibility	133
5.7	Conclusions	136
Chapter 6: Non-Gaussian information from weak lensing data via deep learning		139
6.1	Introduction	139
6.2	Data	142
6.2.1	Mock convergence maps	142
6.2.2	Neural network training and architecture	144
6.2.3	Alternative descriptors	149
6.3	Results	152
6.4	Discussion	157
6.4.1	Non-Gaussian information extracted by the neural network	157
6.4.2	Effect of the smoothing scale on the results	160
6.4.3	Bias in the CNN predictions	163
6.5	Conclusions	165
Chapter 7: Interpreting deep learning models for weak lensing		167
7.1	Introduction	167
7.2	Model and data	169

7.3	Network performance relative to alternative statistics . . . . .	170
7.4	Interpreting DNNs with saliency methods . . . . .	175
7.4.1	Method comparison and selection . . . . .	177
7.4.2	Mapping attributions back to physical space . . . . .	180
7.5	Discussion and conclusions . . . . .	182
Chapter 8: Conclusions and future work . . . . .		184
8.1	Summary of results . . . . .	184
8.2	Future work . . . . .	186
References . . . . .		212
Appendix A: Gaussian likelihood approximation . . . . .		213
Appendix B: Sensitivity of results to interpolation . . . . .		216
Appendix C: Impact of filter misalignment and centering errors. . . . .		218
Appendix D: Variance of aperture filter dipole measurements . . . . .		221
Appendix E: Selecting galaxies from MaNGA for stacking. . . . .		223

## List of Tables

2.1	<i>Parameters of the eight models explored around the fiducial model (<math>\Omega_m = 0.26</math>, <math>w = -1.0</math>). All models are spatially flat with <math>\Omega_\Lambda = 1 - \Omega_m</math> and consider a constant equation of state parameter <math>w</math> for DE.</i>	23
2.2	<i><math>\Delta\chi^2</math> for different cosmological models computed for the power spectrum and three non-Gaussian observables (equilateral bispectrum, peak counts and Minkowski functionals) over noisy <math>\kappa</math> maps with source galaxies at either <math>z = 1</math> or <math>z = 2</math>.</i>	37
2.3	<i>Marginalized errors on <math>\Omega_m</math> and <math>w</math>, orientation of the Fisher ellipse (measured as the angle between its major axis and the <math>w</math> axis), and figure-of-merit (FOM; defined as <math>\pi/A</math>, with <math>A</math> the area of the error ellipse). The errors correspond to a 68% confidence level, scaled to a <math>1000 \text{ deg}^2</math> survey. All calculations were done on noisy <math>\kappa</math> maps with source galaxies at either <math>z = 1</math> or <math>z = 2</math>.</i>	39
3.1	<i>Cosmological parameters for the fiducial model. All other cosmologies share these parameters except <math>\Omega_m</math> and <math>\sigma_8</math>.</i>	50
3.2	<i>The main tunable parameters of CAMELUS and their values used in this study. Dark matter halos are assumed to have a Navarro-Frenk-White (NFW) density profile, defined by its inner slope (<math>\alpha</math>), and its concentration parameter, the ratio between the virial and scale radii, determined itself by <math>c_0</math> and <math>\beta</math>: <math>c_{\text{NFW}} \equiv \frac{c_0}{1+z} \left( \frac{M}{M_\star} \right)^\beta</math>, where <math>z</math> is the halo's redshift, <math>M</math> its mass and <math>M_\star</math> its pivot mass (see [170] for a detailed description of the halo density profile characterization).</i>	53
3.3	<i>Description of the thresholds used in this study to bin the convergence peak counts by their signal-to-noise (<math>S/N</math>) ratio, as well as the mean peak counts from data obtained from both the N-body and the CAMELUS models in the fiducial cosmology, in the bins used for inference.</i>	56

3.4	Comparison of $L_{cg}$ $2\sigma$ (95.4%) credible contours. The figure-of-merit, $FoM$ , is the inverse of the area of the credibility regions. Also displayed are the percentage changes in the area of the credibility region and its centroid (arithmetic mean) shift. We find looser constraints ( $\approx 30\%$ ) for N-body data, whose predictive power is greatly diminished when low significance peaks are excluded from the analysis.	63
3.5	Effect of using a cosmology-dependent covariance matrix. $1\sigma$ (68.3%) and $2\sigma$ (95.4%) credible contours are computed using the three likelihoods described in § 3.2.3 ( $L_{cg}$ , $L_{svg}$ and $L_{vg}$ ). The analysis is done twice, using only high significance peaks ( $S/N > 3$ ) and all the peaks. We report the figure of merit (FoM); defined as the inverse of the area of the credibility region), changes in the credibility regions and shifts in their centroid. Introducing a cosmology-dependent covariance into the $\chi^2$ term of the Gaussian likelihood has a bigger impact than introducing it in the determinant term. Also, the effect is bigger when only high peaks are included.	64
3.6	Impact on the models' predictive power of the lowest significance peak bin included in analysis. Figure of merit (FoM) and change in $2\sigma$ contour area ( $\Delta\%$ ) for constant, semi-varying and variable covariance likelihoods.	70
4.1	Memory and computational time requirements for main simulation tasks for different hyper-parameter values. Each CPU time unit is a core hour (representing wall clock time if computed in series). Changes in CPU time are relative to the fiducial run (in bold).	84
4.2	Individual plane and snapshot size and the respective total storage requirements for both intermediate data products. In bold, values for the fiducial configuration.	85
4.3	Goodness of fit for different lens plane thickness configurations, based on the reduced $\chi^2$ (i.e. $\chi^2$ per degree of freedom). Configurations that yield good fits ( $\chi^2 \leq 1$ ), implying that they are indistinguishable from the fiducial case, are highlighted in bold. <b>Power spectrum:</b> values for a range of $\ell \in [200, 12000]$ and in parenthesis $\ell \in [200, 3532]$ . <b>PDF:</b> values for a range of $\kappa \in [-3.0, 5.0]$ in units of the shape noise r.m.s., and in parenthesis $\kappa \in [-3.0, 3.1]$ . <b>Peak counts:</b> values for a range of $\kappa \in [-0.5, 5.0]$ in units of the shape noise r.m.s., and in parenthesis $\kappa \in [-0.5, 4.0]$ . <b>MFs:</b> values for a range of $\kappa \in [-2.0, 5.0]$ in units of the shape noise r.m.s., and in parenthesis $\kappa \in [-2.0, 3.0]$ .	97

4.4	Goodness of fit as in Table 4.3, but for different mass resolution configurations. Configurations with $\chi^2 \leq 1$ , implying that they are indistinguishable from the fiducial case, are highlighted in bold. <b>Power spectrum:</b> values for a range of $\ell \in [200, 12000]$ and in parenthesis $\ell \in [200, 3532]$ . <b>PDF:</b> values for a range of $\kappa \in [-3.0, 5.0]$ in units of the shape noise r.m.s., and in parenthesis $\kappa \in [-3.0, 3.1]$ . <b>Peak counts:</b> values for a range of $\kappa \in [-0.5, 5.0]$ in units of the shape noise r.m.s., and in parenthesis $\kappa \in [-0.5, 4.0]$ . <b>MFs:</b> values for a range of $\kappa \in [-2.0, 5.0]$ in units of the shape noise r.m.s., and in parenthesis $\kappa \in [-2.0, 3.0]$ . . . . .	103
5.1	Instrumental configurations considered for different existing and planned CMB experiments, defined in each case by the beam's FWHM and the rms noise level. . .	118
5.2	The number of galaxies required for a $3\sigma$ detection of a rkSZ signal. For each combination of CMB experiment and galaxy survey type, the number on the left corresponds to the aperture filter (measured at $0.1 R_{\text{vir}}$ ) and the one on the right to matched filtering. The combination of <i>Planck</i> and a SAMI-like survey, using the aperture filter and adopting the multi-phase slow-rotator model, yields a required number of galaxies exceeding those available within the stellar-mass and redshift range of the galaxy survey. This detection is therefore impossible, and is marked in italics (see § 5.6.3 for discussion). . . . .	122
5.3	Minimum footprint size of surveys required for a $3\sigma$ detection of a rkSZ signal, in $\text{deg}^2$ . This corresponds to the required number of galaxies, shown in Table 5.2, and converted to sky coverage using the stellar mass function described in § 5.6.3. An "X" indicates that the required sky area exceeds the full sky. The number on the left corresponds to the requirement using an aperture filter (measured at $0.1 R_{\text{vir}}$ ) and the one on the right to the use of matched filtering. . . . .	134
6.1	<i>Summary of the neural network's architecture. Convolutional layers increase the depth of the network by applying different filters (sub-layers) to the same input. The number of neurons in a layer is determined by the dimension of its output. The number of weights for a convolutional layer is given by <math>F_{\text{out}} (F_{\text{in}} \times 9 + 1)</math>, where <math>F_{\text{out}}</math> is the number of feature maps that the layer outputs and <math>F_{\text{in}}</math> the number of feature maps the layer is fed with. A fully connected layer is defined by <math>(N_{\text{in}} + 1) \times N_{\text{out}}</math> weights, where <math>N_{\text{in}}</math> is the number of nodes in the previous layer and <math>N_{\text{out}}</math> the number of nodes in the fully connected layer. . . . .</i>	147
6.2	<i>Standard deviation (<math>\times 10^3</math>) of the predictions for the parameters <math>\{\Omega_m, \sigma_8, \Sigma_8\}</math>, averaged for all the cosmological models. In parenthesis, values for the fiducial model.</i>	154
6.3	<i>Area of the 68% and 95% <math>\{\Omega_m, \sigma_8\}</math> credible contours, relative to those obtained from the output of the neural network for un-smoothed, noiseless <math>\kappa</math> maps. . . . .</i>	157

6.4	<i>Area of the 68% and 95% <math>\{\Omega_m, \sigma_8\}</math> credible contours, relative to those obtained from the output of the neural network, for different smoothing scales of <math>\kappa</math> maps. The first row corresponds to the un-smoothed data.</i>	162
7.1	Percentage change in the area of credible contours derived from different statistics when they are combined with the output from the DNN (see Fig. 7.1 for a graphical representation of those contours). The statistic used for the right-most column (labelled “All”) is a combination of all the statistics in the other columns: power spectrum (PS), lensing peaks (PC), and Minkowski functionals ( $V_0, V_1, V_2$ ). The change in area is defined as $\Delta \text{Area} = 100 \left( \frac{\text{Area}_{w/DNN}}{\text{Area}_{w/o DNN}} - 1 \right)$ .	172
7.2	Correlation measurements between the saliency maps computed from the fitted DNN, and the same architecture with all the parameters randomized (see columns labelled “Trained model” and “All layers randomized” in Fig. 7.3). For all three tests (Pearson, Spearman, and Kendall), the null hypothesis is that there is no relationship between both saliency maps. P-values are two-sided.	180



## List of Figures

1.1	Comparison of parameter constraints from a series of WL analyses of galaxy surveys and <i>Planck</i> . (Figure from [20]). . . . .	6
1.2	Comparison of bias-variance trade-off for traditional machine learning algorithms on the left, and the different regime for DNNs on the right (Figure from [78]). . . .	15
2.1	<i>Comparison between the power spectra measured for selected models, as labeled, over noiseless, un-smoothed <math>\kappa</math> maps (thick lines) and analytic predictions using a fitting formula [146] for the matter power spectrum (thin lines). Percent differences between measured and predicted power spectra are depicted in the lower panel. Shaded areas represent <math>\pm 1</math> standard deviations around the average, scaled to a <math>1000 \text{ deg}^2</math> survey, and in the lower panel only the standard deviation for the fiducial model is plotted for reference.</i> . . . . .	25
2.2	<i>(color online)Sensitivity of the power spectrum to <math>\Omega_m</math> and <math>w</math> for noiseless (upper panels) and noisy (lower panels) convergence. Estimates including only geometry effects are shown in red, those including only growth effects in blue, and those including both effects in black. In the upper panels, analytic predictions are displayed with thin lines, for comparison. Source galaxies are at <math>z_s = 1.0</math> in all cases. Shaded areas represent a <math>\pm 1</math> standard deviation around the measured averages scaled to a survey sky coverage of <math>1000 \text{ deg}^2</math> and only selected models are displayed for clarity.</i> . . . . .	29
2.3	<i>(color online)Sensitivity of the equilateral bispectrum of the noiseless convergence field to <math>\Omega_m</math> and <math>w</math>. Both panels show the percentage deviation in each model from the fiducial bispectrum. For clarity, only two models are depicted per panel, with the source galaxies at <math>z_s = 1</math>. As in Fig. 2.2, black lines show the net sensitivity, red lines the sensitivity due only to differences in geometry and blue lines the sensitivity due only to differences in growth. Shaded areas represent <math>\pm 1</math> standard deviation around the measured averages, scaled to a <math>1000 \text{ deg}^2</math> survey.</i> . . . . .	31

2.4	(color online)Sensitivity of peak counts to $\Omega_m$ and $w$ on noisy convergence maps. Both panels show the percentage difference between the peak counts in a given cosmology and in the fiducial model. Peak height is expressed in units of $\kappa$ and in units of $\sigma_{\text{noise}}$ , $S/N$ . For clarity, only two models are depicted per panel, with source galaxies at $z_s = 1$ . The color scheme is the same as in Figs. 2.2 and 2.3. Shaded areas represent $\pm 1$ standard deviation around the measured averages, scaled to a $1000 \text{ deg}^2$ survey. . . . .	33
2.5	(color online)Percentage difference of the three MFs measured on noisy $\kappa$ maps, compared to the value in the fiducial model, when changing $\Omega_m$ and $w$ . Left/center/right panels show the results for $V_0/V_1/V_2$ , for noisy $\kappa$ and source galaxies at $z_s = 1$ . The color scheme, labeled in the legends, is the same as in Figs. 2.2-2.4. Shaded areas represent $\pm 1$ standard deviation around the measured averages, scaled to a $1000 \text{ deg}^2$ survey. . . . .	35
2.6	(color online)68% Fisher error ellipses in the $(\Omega_m, w)$ plane inferred from the power spectrum ( $P_l$ ), equilateral bispectrum ( $B_{lll}$ ), lensing peaks and Minkowski functionals (MFs). Upper/lower panels show the contours for source galaxies at $z_s = 1/z_s = 2$ . Each observable was characterized by a data vector of length 20, and the ellipses were computed neglecting the cosmology-dependence of the covariance matrix. All contours are scaled to a $1000 \text{ deg}^2$ survey. . . . .	40
2.7	(color online)Sensitivity of lensing peak counts to $\Omega_m$ , derived from a set of Gaussian random fields with the same power spectra as that measured on noisy convergence maps from large-scale structure. Shaded areas represent 1 standard deviation errors in a $1000 \text{ deg}^2$ survey. Compare with the left panel of Fig.2.4. . . . .	42
2.8	(color online)Sensitivity of DM halo abundance to $\Omega_m$ (left panel) and $w$ (right panel). The percentage difference in the total number of halos per unit solid angle to $z = 1$ between a model and the fiducial cosmology, as a function of the halo mass. The net effect (black) is decomposed into its geometry (red) and growth (blue) components. . . . .	44
2.9	(color online)Difference in number of peaks from the fiducial cosmology, normalized by the standard deviation in the fiducial model, for a $1000 \text{ deg}^2$ survey. . . . .	45
2.10	(color online)Difference $\Delta\kappa$ between a single realization of the noisy $\kappa$ field generated including only geometry (top panel) or growth effects (bottom panel) for $w = -1.2$ , and the corresponding realization in the fiducial model ( $w = -1.0$ ). . . . .	46

3.1	Comparison of mean peak counts as a function of their height between N-body simulations (blue) and CAMELUS (red). Counts are normalized to 1 deg <sup>2</sup> of sky and height is expressed in absolute value and as a signal-to-noise ratio ( $S/N$ ). The upper panels show the results from smoothed convergence maps without shape noise; the lower panels add shape noise. Three different cosmologies are displayed with increasing parameter $\Sigma_8$ from left to right (0.556, 0.734 and 1.244). In black, we show the fractional difference between the two models ( $\Delta[\%] \equiv (N_{\text{CAMELUS}} - N_{\text{N-body}})/N_{\text{N-body}}$ ), and the area between the histograms is shaded. Adding noise reduces the discrepancies between the models but the effect depends on cosmology. While the discrepancies are almost erased for cosmologies with small $\Sigma_8$ , for the rest N-body data yield lower counts near the maximum of the distribution and higher counts in the tails. The differences grow with $\Sigma_8$ . . . . .	58
3.2	Global comparison of peak counts. For each cosmology, the area between the N-body and CAMELUS histograms as a percentage of the area enclosed by the N-body histogram (Eq. 3.17) is plotted against $\Sigma_8$ . Differences from noiseless maps (crosses) are significantly reduced by adding noise (dots), so that the difference stays below 20% in all cases. The reduction is more important for cosmologies with small $\Sigma_8$ , for which noise dominates. . . . .	59
3.3	Covariance comparison between N-body (upper panel) and CAMELUS (lower panel) for the fiducial cosmology. Each normalized covariance matrix has diagonal elements equal to the peak count variance divided by its mean, $\frac{\sigma_{ii}^2}{\bar{x}_{ii}}$ , and off-diagonal elements equal to the correlation coefficients, $\rho_{ij} \equiv \frac{\sigma_{ij}}{\sigma_i \sigma_j}$ . We find higher absolute values for all elements in the matrices, with positive and negative peaks positively correlated and positive peaks anti-correlated with negative ones. Peak counts from CAMELUS are mildly anti-correlated. Selected matrix elements whose value for all cosmologies is displayed in Fig. 3.4 are indicated with a number. . . . .	61
3.4	The cosmology dependence of covariances. Each subplot shows the value of selected normalized covariance matrix elements for all 162 cosmologies. The selected elements are indicated in Fig. 3.3, and correspond to a diagonal element (left panel) and off-diagonal elements showing anti-correlation in N-body data (center panel) and correlation (right panel). N-body data exhibit higher absolute values for all elements and stronger cosmology dependence. . . . .	62

- 3.5 Comparison of  $2\sigma$  (95.4%) credible contours from N-body (blue) and CAMELUS (red) data, using a Gaussian likelihood with constant covariance,  $L_{cg}$ . Solid lines show the contours computed using all the peak counts. We find looser constraints, with a thicker,  $\approx 30\%$  larger credibility region. Dashed lines show the results including only high significance peaks ( $S/N > 3$ ). While constraints based on CAMELUS data do not change, the predictive power from N-body data is severely reduced, with a  $\approx 200\%$  increase in the area of the credibility region. Dotted lines show the degeneracies  $\Sigma_8 = \sigma_8 \left( \frac{\Omega_m}{0.3} \right)^\alpha$  that minimize scatter in  $L_{cg}$ . We find a steeper contour,  $\alpha = 0.67$  vs.  $\alpha = 0.58$  for CAMELUS. Grey dots show the simulated cosmologies (a green star the fiducial cosmology), and grey areas the regions excluded from contour measurements. . . . . 63
- 3.6 Effect on the credibility regions of using a cosmology-dependent covariance. In the left panels we show the change caused by introducing a variable covariance in the  $\chi^2$  term of a Gaussian likelihood ( $L_{svg}$ , shaded areas) compared with a constant covariance ( $L_{cg}$ , lines). On the right we display the change from using a variable covariance matrix in the determinant term as well ( $L_{vg}$ , shaded areas) compared with  $L_{svg}$  (lines). The upper panels show the result using only high-significance ( $S/N > 3$ ) peaks, while the lower panels show results with all peaks included. Introducing a variable covariance in the  $\chi^2$  has a larger impact than using it in the determinant term. Also, the effects are larger when using only high significance peaks (see Table 3.5). . . . . 65
- 3.7 Impact of differences in peak counts and covariance matrices on credible contours. Solid lines are  $2\sigma$  contours from N-body (blue) and CAMELUS (red) data. Magenta lines are contours computed mixing peak counts from one model with the covariance matrices from the other. The dashed contours combine N-body covariance matrices with CAMELUS peak counts, and the dotted contours combine conversely N-body peak counts with CAMELUS covariances. The upper panels show the results using only  $S/N > 3$  peaks while the lower panels display the contours obtained including all peaks. On the left we show contours computed using a constant covariance,  $L_{cg}$  and on the right those introducing a variable covariance in the  $\chi^2$  term,  $L_{svg}$ . In general, contours computed with the same covariance matrices are closer than those calculated with the same peak counts. The effect is more noticeable for the cases which include only high-significance peaks, since for these the N-body and CAMELUS contours exhibit a greater difference. . . . . 67

3.8	Peak and halo count variance comparison between N-body (blue) and CAMELUS (red). <b>Upper panel:</b> ratio of the cumulative peak count standard deviation from its value expected for a Poisson distribution, as a function of peak height. For pure Poisson shot noise, this ratio is unity (horizontal black dashed line). We find significantly higher sample variance than the results from CAMELUS, and what would be expected for a Poisson distribution. As the peak height increases and the peak counts decrease, shot noise starts to dominate. <b>Lower panel:</b> variance of the cumulative halo number as a function of minimum halo mass. Sample variance is estimated from different sub-volumes, and scaled to a common reference volume of $(250 h^{-1}\text{Mpc})^3$ . Shot noise is estimated from a Poisson distribution with mean value adopted from a theoretical halo mass function [172]. N-body cumulative halo counts exhibit a sample variance higher than expected from a Poisson distribution. Shot noise becomes more important at higher masses, as the halos become scarcer. CAMELUS is dominated by shot noise. . . . .	69
3.9	Influence on the credibility region of the lowest significant peaks included in the ( $L_{cg}$ ) likelihood calculation. <b>Upper panel:</b> for N-body simulations, including peaks with $2.0 < S/N < 3.0$ significantly improves the model's predictive power. <b>Lower panel:</b> for CAMELUS, little or no improvement in predictive power is found when lower-significance peaks are included. . . . .	71
3.10	Effect on CAMELUS credible contours of finite sampling of the cosmological parameter space. $2\sigma$ contours obtained from a fine grid of 7,803 models (black) and interpolated from our suite of 162 cosmologies (red). The interpolated contour is smaller in the low- and high- $\Omega_m$ tails. Thus we excluded from our analyses the greyed-out regions, corresponding to $\Omega_m < 0.160$ and $\Omega_m > 0.600$ . . . . .	72
3.11	Effect of cosmological parameter sampling on the N-body credible contours. We draw 1,000 bootstrap samples from our suite of cosmologies; i.e., we draw samples of 162 elements with substitution, each having on average 102-103 different cosmologies. <b>Upper panel:</b> $L_{cg}$ $2\sigma$ contours from the full suite (black) and the 1,000 sub-samples (blue). Darker areas indicate higher contour concentration. <b>Lower panel:</b> area histogram for the bootstrap samples. Displayed for reference are the area for the full suite (black dashed line), 90% of this value (grey dashed line) and 80% (grey dotted line). 81% of the contours fall within 10% of the original area and 99% within 20%. . . . .	73

4.1	Comparison of convergence power spectra measured in our simulations with $1024^3$ DM particles, to those obtained in semi-analytic calculations using the flat-sky approximation and <code>halofit</code> . <b>Upper panel:</b> Convergence power spectra. Thick, black lines correspond to simulations, while thin, blue lines correspond to the <code>halofit</code> calculations. Different dashes indicate different redshift bins for the lensed galaxies. Full lines correspond to results in the presence of shape noise (see section § 4.2.1 for a description of the shape noise level considered), and partly transparent lines to the results in the absence of noise. <b>Lower panel:</b> percentage difference between simulations and <code>halofit</code> , for each redshift bin. As in the upper panel, full lines correspond to the noisy, and partly transparent lines to the idealized noiseless case. . . . .	80
4.2	Effect of lens plane thickness on convergence power spectra. Each column shows, for a different redshift bin, the percentage difference in the mean auto power spectrum measured over 10,048 convergence maps for a given thickness of the lens planes, relative to the fiducial value of $80 h^{-1}$ Mpc. The top row corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines represent measurements over simulated maps, while thin lines represent theoretical predictions following [201]. For comparison, a standard error is shown in shaded gray, corresponding to 3 standard deviations of the measurements in the fiducial case, scaled to a $2 \times 10^4 \text{ deg}^2$ survey. . . . .	90
4.3	Same as Figure 4.2 for the one-point $\kappa$ probability density function (PDF). Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps. . . .	91
4.4	Same as Figure 4.2 for the lensing peak distribution. Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps. . . . .	92
4.5	Same as Figure 4.2 for the three Minkowski functionals (MFs). Each pair of rows shows a different MF. The top row of each pair corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps. . . . .	94
4.6	Covariance matrix measured over 10048 convergence maps generated from our $512^3$ particle simulation with $80 h^{-1}$ Mpc lens planes. The axis labels indicate the redshift bins used. For example, the slice $[(1.0, 1.5), (2.5, 2.5)]$ shows the correlation between the cross power spectrum of redshift bins $z_s = 1.0$ , $z_s = 1.5$ and the auto power spectrum of the redshift bin $z_s = 2.5$ . . . . .	95

4.7	Effect of mass resolution (number of particles in the simulation volume) on the convergence power spectrum. Each column shows, for a different redshift bin, the percentage difference in the mean auto power spectrum measured over 10,048 convergence maps for a given mass resolution of the underlying N-body simulation, relative to the fiducial value of $1.1 \times 10^{10} M_{\odot}$ (corresponding to $1024^3$ DM particles in the simulation box). The top row corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines represent measurements over simulated maps, while thin lines represent theoretical predictions following Eq. (4.5). For comparison, a standard error is shown in shaded gray, corresponding to 3 standard deviations of the measurements in the fiducial case, scaled to a $2 \times 10^4 \text{ deg}^2$ survey. . . . .	98
4.8	Same as Figure 4.7 for the one-point $\kappa$ probability density function (PDF). Thick lines represent measurements over simulated maps, while thin lines represent the expectation for GRFs with the same power spectrum as the $\kappa$ maps. . . . .	100
4.9	Same as Figure 4.2 for the lensing peak distribution. Thick lines represent measurements over simulated maps, while thin lines represent the expectation for GRFs with the same power spectrum as the $\kappa$ maps. . . . .	100
4.10	Same as Figure 4.2 for the three Minkowski functionals (MFs). Each pair of rows shows a different MF. The top row of each pair corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps. . . . .	102
5.1	Density profiles for the ionized gas for fully ionized galactic atmospheres (“hot”, in red) and multi-phase atmospheres (“multi-phase”, in blue) as a function of distance to the center in virial radii units for galaxies of three different stellar masses. The gas metallicity is $Z_g = 0.3 Z_{\odot}$ and the DM density profile of the host halo is shown in black for reference. The halos’ virial masses are $\{1.2 \times 10^{11}, 3.4 \times 10^{11}, 3.3 \times 10^{12}\} M_{\odot}$ , their virial radii $\{69, 98, 215\}$ kpc, and their concentrations $\{17.3, 15.0, 11.0\}$ , respectively. . . . .	110
5.2	Radial profiles of the tangential gas velocity for a fast rotator (in red, rotating at the halo’s circular velocity) and a slow rotator (in blue, rotating at a velocity consistent with simulations [235]) for galaxies of three different stellar masses. . . . .	112

- 5.3 **Upper panel:** 2D map of the expected fractional temperature change induced in the CMB by the rotation of a  $M_* = 10^{10} M_\odot$  galaxy hosted by a  $3.4 \times 10^{11} M_\odot$  DM halo with a virial radius of 98.1 kpc and a concentration parameter of  $C_v = 15.0$  at a redshift of  $z = 0.03$ , assuming the multi-phase slow model. The galaxy's inclination is 1 rad and the signal has been convolved with a 5 arcmin FWHM beam (represented by the small dotted circles at the center), while the halo virial radius has an apparent size of 5.2 arcmin. **Lower panel:** cut along the X-axis of the dipole-like signal in the upper panel, for galaxies of three different stellar masses. The predictions in the hot+fast model (fully ionized atmosphere rotating at the halo's circular velocity) are shown in red, and the multi-phase, slow model are shown in blue. The signal was calculated for a metallicity of  $Z_g = 0.3 Z_\odot$ , and shown in absolute value. The profile of the convolving beam is displayed in black. 113
- 5.4 The distribution of galaxies in redshift and stellar mass, in the two prototype surveys considered to assess the feasibility of detecting the kSZ signal induced by the rotation of galactic halos. The top panel corresponds to the primary sample in MaNGA, and the lower panel to SAMI. . . . . 117
- 5.5 Comparison between the power spectrum of CMB temperature anisotropies, including instrumental noise which dominates at  $\ell \gtrsim \text{few} \times 10^3$ , and that for the expected rkSZ signal for three galaxies of different stellar mass (the same galaxies used in Figs. 5.1-5.3). The scale on the left y-axis refers to the CMB+noise and the scale on the right to the kSZ power spectra. Note that the rkSZ power is several orders of magnitude lower than that from the combination of CMB + instrumental noise. . . . . 119
- 5.6 Number of galaxies needed for a  $3\sigma$  detection of the rkSZ signal using an aperture filter, as a function of filter size. Each CMB experiment configuration is displayed in a different color. The galaxy  $(z, M_*)$  distribution is assumed to be that of the primary sample of MaNGA. Solid lines correspond to predictions based on the hot+fast model, while dashed lines are based on the multi-phase, slow model. For reference, vertical lines represent the CMB experiment beam size, in units of the average  $R_{\text{vir}}$  for the galaxies in the survey. . . . . 121
- 5.7 *Planck* SMICA-noSZ CMB data stacked on the positions of MaNGA galaxies (with equal weight) after aligning with the galaxies' spin angles, and scaling to their  $R_{\text{vir}}$ . The left panel shows the results for all 2,664 galaxies in our sample, the central panel for the 1,953 spiral galaxies, and the right panel for the 1,235 field spiral galaxies. For each galaxy, located at the origin (0,0), we stacked spin-aligned  $8 \times 8 R_{\text{vir}}$  patches. The circles correspond to  $R_{\text{vir}}$  and the vertical lines mark the expected galaxy spin direction (pointing towards the -y axis, the right half is approaching the observer, and the left side receding). Any rotation-induced temperature dipole should show a left-right cold-hot temperature asymmetry (see Fig. 5.3). In each text box, the measured dipole on a  $1.0 R_{\text{vir}}$  aperture, and its signal-to-noise ratio based on the theoretical noise calculation described in 5.3.1. . 127



5.8	Probability distribution function (PDF) of the temperature dipole of the SMICA-noSZ map stacked on the positions of 1,253 field spirals after randomizing their spin angles. The PDF is inferred from $10^4$ measurements using $1.0 R_{\text{vir}}$ aperture filters that have been randomly placed and rotated. Superimposed, a Gaussian PDF with the same mean and standard deviation as the $10^4$ measurements, shows good agreement with the data. Dotted vertical lines correspond to $1\sigma$ , $2\sigma$ , and $3\sigma$ thresholds. The solid vertical line is the measured dipole on the stack with the galaxies' spins aligned. . . . .	128
5.9	Single-frequency <i>Planck</i> CMB maps from its high-frequency instrument (HFI), stacked with equal weights on the positions of the 1,235 field spirals after aligning their spin angles. Note the difference in units for the two highest-frequency maps: MJy/sr instead of K. As with Fig. 5.7, the circle represents the average virial radius and the vertical line the spin direction. The text boxes show the measured dipole, as well as its percentile and S/N ratio. The percentile and significance have been derived from noise-only stacks. . . . .	129
5.10	The figure illustrates the impact of a factor-of-two reduction in the total baryon mass in the halo relative to the cosmic mean value. The corresponding reduction in the density of the hot, ionized gas component is shown as a function of radius, in units of the virial radius. Results are shown for the same three halos as in the previous figures, labelled by the stellar mass of their central galaxies. In the hot model, where all the baryons are ionized, the gas density scale linearly with the baryon fraction, independent of mass and radius (shown in red). The hot coronae of the multi-phase gaseous halos (shown in blue) are less affected, except in the outer regions where the rKSZ signal contributes little to the SNR. The effect on the multi-phase model decreases with smaller halo mass. . . . .	131
6.1	<i>Location of the 96 cosmological models in our dataset on the <math>\{\Omega_m, \sigma_8\}</math> plane. The fiducial model, <math>\{\Omega_m = 0.260, \sigma_8 = 0.800\}</math>, is marked by a red star, and grey lines delimitate the quadrants defined by the fiducial parameters. The quadrants labeled <b>I</b> and <b>II</b> are discussed in § 6.4.3. The dashed curves show isolines for <math>\Sigma_8 \equiv \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^{0.6}</math> for reference. . . . .</i>	143
6.2	<i>Schematic representation of the convolutional neural network (CNN) used in this study. The network consists of a series of convolutional and (average) pooling layers. Layers increase their “logical” dimension (depth), while reducing their “spatial” dimensions (width and height). Once the spatial dimension has been reduced to unity (flattening), a series of fully connected layers further reduces the number of nodes to two, the required number of outputs. The activation function for the neurons is a leaky rectified linear unit. For clarity, only a few layers are displayed. . . . .</i>	146

6.3	<i>Comparison of the average convergence power spectrum for the fiducial <math>\kappa</math> maps with predictions from linear and non-linear theory. The theoretical curves were computed using NICA EA [278], with the revised Halofit parameters from [195] for the non-linear power spectrum. . . . .</i>	150
6.4	<i>Comparison of peak counts derived from maps generated via our ray-tracing N-body simulations, to those derived from Gaussian random fields (GRFs) with the same power spectrum. . . . .</i>	151
6.5	<i>Predictions for <math>\{\Omega_m, \sigma_8, \Sigma_8\}</math> from un-smoothed (<math>\approx 0.2</math> arcmin/pixel) convergence maps, compared to their true values. Each point represents a map in the test data set. Predictions from the CNN are displayed in red, from the power spectrum in blue and from peak counts in green. Vertical dashed lines indicate the true values for the fiducial cosmology, and diagonal dashed lines the unbiased Prediction = Truth relationship. The dashed rectangles in the middle and right panels mark a small set of realizations of models near the fiducial cosmology; these contain anomalous structures leading to large biases (see text for discussion). . . . .</i>	153
6.6	<i>68% and 95% credible contours for un-smoothed (<math>\approx 0.2</math> arcmin/pixel) <math>\kappa</math> maps, derived from the power spectrum (blue), lensing peak counts (green) and neural network predictions (red). The true values for the parameters, <math>\{\Omega_m = 0.260, \sigma_8 = 0.800\}</math> are indicated by black dotted lines. The upper and right panels show the distribution marginalized over the other parameter. . . . .</i>	156
6.7	<i>Same as Fig. 6.6, except using the Gaussian random fields, rather than the ray-tracing simulations. The network was trained with the un-weighted loss function (eq. 6.2). . . . .</i>	159
6.8	<i>Same as Fig. 6.5, except using the Gaussian random fields, rather than the ray-tracing simulations. The network was trained with the un-weighted loss function (eq. 6.2). . . . .</i>	160
6.9	<i>Same as Fig. 6.6, except smoothing the maps from the ray-tracing simulations with a Gaussian kernel of 1 arcmin of width. The network was trained with the un-weighted loss function (eq. 6.2). . . . .</i>	161
6.10	<i>Predictions from the CNN for <math>\{\Omega_m, \sigma_8, \Sigma_8\}</math> from unsmoothed (<math>\approx 0.2</math> arcmin/pixel) convergence maps, compared to their true values. Each point represents a map in the test data set. Predictions using the unweighted loss function (eq. 6.2) are displayed in grey, and those using a weighted loss function (eq. 6.3), to account for the heterogeneous sampling of the parameter space, in red. Vertical dashed lines indicate the true values for the fiducial cosmology, and diagonal dashed lines the unbiased Prediction = Truth relationship. . . . .</i>	163

6.11	68% and 95% credible contours for un-smoothed ( $\approx 0.2$ arcmin/pixel) $\kappa$ maps, derived from two neural networks with the same architecture: in red the result from training with the weighed loss function (eq. 6.3) and in grey the result from training with the un-weighted loss function (eq. 6.2). True values are indicated by black, dotted lines. The upper and right panels show the marginal distribution for $\Omega_m$ and $\sigma_8$ , respectively. . . . .	164
7.1	Credible contours derived for $\Omega_m$ and $\sigma_8$ . Each panel shows the comparison between the constraints derived from the DNN (in red) from an alternative statistic (in blue), and the combination of the DNN and the statistic (in black). Solid lines enclose 68% of the likelihood, and dot-dashed lines 95%. Upper row, from left to right: comparison between the DNN and a combination of statistics, the power spectrum (PS), and lensing peak counts (PC). Lower row, from left to right: comparison between the DNN and the three Minkowski functionals, $V_0$ , $V_1$ , and $V_2$ . The grey dots indicate the points in parameter space for which simulations were available. . . . .	173
7.2	Pearson's correlation coefficient (averaged over the 101 cosmologies) between the DNN predictions for $\Omega_m$ (solid lines) and $\sigma_8$ (dotted lines), and the measured statistics. For the power spectrum (PS, blue), the bins correspond to different multipoles ( $\ell$ ; see upper scale), and for the other statistics, values of $\kappa$ (lower scale). . . . .	174
7.3	Examples of saliency maps for the output neuron of the DNN that encodes the parameter $\Omega_m$ . The left-most column ("Input") shows a small region ( $100 \times 100$ pixels, or $0.68 \times 0.68$ deg <sup>2</sup> ) of a $3.5 \times 3.5$ deg <sup>2</sup> $\kappa$ map from the fiducial cosmology. The second column ("Trained model") shows the region of the saliency maps that corresponds to the region of the input map on the left. The third column ("Last layer randomized") shows the same saliency map as the second column, computed on the fully trained model after randomizing the weights of the last (output) layer. The right-most column ("All layers randomized") shows the same saliency map as columns 2-3, computed on a model where all the weights are randomized. Each row corresponds to a different saliency method. The scales for each image are omitted for clarity, since they do not influence the conclusions. . . . .	179
7.4	<b>Upper panels:</b> sum of the square of the pixel values in saliency maps as a function of $\kappa$ in the corresponding input maps. Each line is the test maps' average for one of the 101 cosmologies. <b>Lower panels:</b> same as upper panels, divided by the number of pixels in each $\kappa$ bin, giving the mean saliency <sup>2</sup> per pixel as a function of $\kappa$ . Left panels correspond to saliency maps computed using the <i>Gradient</i> method, and right panels to saliency maps computed using the <i>Input<math>\times</math>gradient</i> method. Each line is colored based on the value of $S_8 = \sigma_8 \left( \frac{\Omega_m}{0.3} \right)^{0.6}$ . . . . .	181

A.1	<i>Credible contours for <math>\{\Omega_m, \sigma_8, \Sigma_8\}</math> from lensing peak counts on noisy <math>\kappa</math> maps. Filled contours correspond to a Gaussian likelihood, and solid lines to contours corresponding to KDE estimates. . . . .</i>	214
B.1	<i>68% and 95% credible contours for un-smoothed (<math>\approx 0.2</math> arcmin/pixel) <math>\kappa</math> maps, derived from two neural networks with the same architecture: the original one trained on all 96 cosmologies (red) and another one for which the model <math>\{\Omega_m = 0.261, \sigma_8 = 0.802\}</math> was excluded (grey). The assumed true value (<math>\{\Omega_m = 0.261, \sigma_8 = 0.802\}</math>) is indicated by black dotted lines. The upper and right panels show the marginal distribution for <math>\Omega_m</math> and <math>\sigma_8</math>, respectively. . . . .</i>	217
C.1	<i>Mean suppression in the measured rkSZ signal due to decentering relative to the halo position and to misalignment relative to the galaxy's projected spin angle. Color indicates whether the rkSZ signal corresponds to that of a hot, fast or cold, slow rotator. The intensity of the color indicates a different CMB experiment configuration (<i>Planck</i> for strong color, CMB-S4 for the faintest color and ACT for the intermediate intensity). Finally, the type of line used indicates the mass of the galaxy. <b>Upper panels:</b> Effect of a Gaussian error in the position of the filter, relative to the halo's center, as a function of the error's FWHM in units of the virial radius. On the left, effect for an aperture filter measuring the signal's dipole. On the right, effect for the measured correlation of a matched filter and the signal at the estimated (erroneous) halo center. <b>Lower panels:</b> Effect of a Gaussian error in the orientation of the filter axis relative to the galaxy's projected spin vector, as a function of the error's FWHM in degrees. As in the upper panels, on the left the effect for the aperture filter is displayed and on the right, for the matched filter. . . .</i>	220
E.1	<i>Histogram showing the difference between the spin angle computed using the <math>H_\alpha</math> line and the angle derived from O II (average of both O II lines). A mixture model with a Gaussian and a uniform component is a poor fit (indicated by the fat tails). To compensate for the badness of fit, we apply a <math>5\sigma</math> cut to identify outliers. . . .</i>	224

## Acknowledgements

The list of people who have helped and supported me during this six year journey is so long that I will almost certainly forget to explicitly mention some, I apologize to them. First I would like to thank my advisor Zoltán Haiman for his guidance and support. I see my PhD as an apprenticeship, and I could not have wished for someone better to learn how to be a scientist. I thank my collaborators, in particular Daniel Hsu for teaching me the little statistics I know, Andrea Petri for making the learning curve to run simulations less steep, Jia Liu for being always willing to work together, and Colin Hill because he taught something new every time we talked. Without the hard work and enthusiasm of Stefan Waterval, Arushi Gupta, and Manasi Sharma, chapters 4, 6, and 7 of this Thesis would not exist.

I thank my first year advisor David Schiminovich for letting me be part of the FireBall-2 team. I have really enjoyed our many discussions of wide-ranging topics, and our time together observing with students. Talking about FireBall-2, I cannot forget all the good moments with Nicole Melso, Julia Blue Bird, Erika Hamden, Michele Limon, and Hwei Ru Ong.

I would like to thank Greg Bryan, David Helfand, and Lam Hui for their continuous advice as members of my thesis committee, and the whole Astronomy Department at Columbia University for making me feel at home. Thank you Marcel Agüeros, Summer Ash, and Jana Grcevich for helping me share my passion for astronomy with our extended community. I thank Mildred Garcia and Ayouné Payne for making my life easier, and my friends and office mates for their patience: Andrea Derdzinski, Adrian Lucy, Emily Sandford, Alejandro Núñez, Maria Charisi, Aleksey Generozov, Mihir Kulkarni, Steven Mohammed, Andrew Emerick, and Shifra Mandel.

Special thanks to my family: Kyla, Addison, and Jocelyn. You have changed my world so much that I cannot imagine life without you. I have never been happier, you help me know myself better, and want to be the best version of me.

Finalmente, quiero dar las gracias a mis padres y hermanas. Soy como soy en gran parte por vosotros, y no habría podido hacer lo que he hecho sin vuestro apoyo incondicional. A pesar de la distancia, siempre estaremos juntos.

*To those who look up at night. And wonder.*

# Chapter 1: Introduction

## 1.1 Weak lensing cosmology

The main unsolved puzzles in cosmology are the nature of dark matter (DM), dark energy (DE), and whether the universe experienced an early inflationary phase. Dark energy is responsible for the accelerated expansion of the universe [1, 2] and constitutes 70% of its energy budget, yet it is the least understood of its components.

Weak gravitational lensing (WL) is the slight deflection of photons emitted at cosmological distances by the inhomogeneous matter distribution along the line of sight. Its main effects are a (de)magnification of the apparent brightness and a small distortion (shear) of the apparent shape of distant galaxies and the anisotropies of the cosmic microwave background (CMB). By statistically measuring these signals, the evolution of the matter density field, both DM and ordinary (baryonic) matter, can be mapped, yielding constraints on the background cosmological model. DE affects the evolution of that field, suppressing the growth of inhomogeneities in the universe and altering how distances between objects grow with time. As a result, the NSF-NASA-DOE joint Dark Energy Task Force identified WL as one of the key probes of DE, and also the one with the highest projected statistical figure of merit [3]. While lensing measurements are dominated by observational effects unrelated to WL, both galaxy and CMB lensing have recently been measured, and WL has become a mature technique.

In this chapter we review the basics of weak lensing, some recent experimental results and a description of upcoming surveys, make the case for the need to access (and use) non-Gaussian information to fully realize the potential of these future experiments, and show how deep learning methods borrowed from other disciplines are finding their way in the analysis of WL data sets. We conclude by outlining the overall structure of this Thesis.

### 1.1.1 Weak lensing fundamentals

In the limit of small deflections and lensed objects (sources) whose apparent size is small compared to the angular scales on which the deflection angle changes (conditions met for galaxy lensing by large scale structure, also known as cosmic shear), the effect of WL corresponds to a mapping of the sources' light distribution onto an image plane. To leading order, this mapping is described by the Jacobian [4, 5]:

$$A_{ij}(\boldsymbol{\theta}) = \frac{\partial \beta_i(\boldsymbol{\theta})}{\partial \theta_j} = \delta_{ij} + \Psi_{ij} \quad (1.1)$$

where a given point  $\boldsymbol{\theta}$  in the image plane corresponds to a true direction  $\boldsymbol{\beta}$  in the source plane, and  $\delta_{ij}$  is the identity matrix. In the weak lensing regime, the distortion matrix  $\Psi_{ij}$ , which defines lensing, is small and can be parametrized in the following way:

$$\Psi_{ij} = \begin{bmatrix} \kappa(\boldsymbol{\theta}) + \gamma_1(\boldsymbol{\theta}) & \gamma_2(\boldsymbol{\theta}) \\ \gamma_2(\boldsymbol{\theta}) & \kappa(\boldsymbol{\theta}) - \gamma_1(\boldsymbol{\theta}) \end{bmatrix} \quad (1.2)$$

The trace-less part of the distortion matrix  $\gamma_{ij}$  represents a shear of the source light distribution, and the convergence,  $\kappa$ , a (de)magnification of the source.

If the luminosity of a population of objects is known, their magnification can be used as a direct estimation of the convergence (see [6] for an application example to survey data). While still challenging, measuring galaxy shapes is easier than knowing their luminosity with high accuracy. In the absence of intrinsic alignments between galaxies, their orientation is random, and correlations in their orientation indicate the presence of lensing. Using the moments of the source brightness distribution,  $I_{obs}(\boldsymbol{\theta})$  to measure their ellipticity:

$$\epsilon_1 \equiv \frac{q_{11} - q_{22}}{q_{11} + q_{22}}, \epsilon_2 \equiv \frac{2q_{12}}{q_{11} + q_{22}} \quad (1.3)$$



with

$$q_{ij} \equiv \int d\boldsymbol{\theta} \theta_i \theta_j I_{obs}(\boldsymbol{\theta}) \quad (1.4)$$

it can be shown (see [7] for the full derivation) that the measured ellipticity over an ensemble of galaxies is an estimator for the reduced shear,  $\mathbf{g}$ , which in the WL limit is approximately equal to the shear.

$$\epsilon_i = \frac{\gamma_i}{1 - \kappa} \frac{1}{1 - \frac{\gamma_1^2 + \gamma_2^2}{(1 - \kappa)^2}} \approx \frac{\gamma_i}{1 - \kappa} \equiv \mathbf{g} \quad (1.5)$$

WL surveys are therefore capable of generating shear maps  $\gamma_i(\boldsymbol{\theta})$  by measuring the ellipticities of many galaxies. In practice, going from pixel intensities in galaxy images to a shear map requires taking into account a plethora of effects, from the instrument's point spread function, to de-blending of source galaxy images, to the effect of intrinsic alignments between galaxies. For a review covering some of these aspects see [8].

The distortion matrix is the Hessian of the lensing potential (projected gravitational potential)

$$\Psi_{ij} = \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j} \quad (1.6)$$

and the shear and convergence can be expressed in terms of the second derivatives of the lensing potential

$$\kappa = \frac{1}{2} \left( \frac{\partial^2 \psi_1}{\partial^2 \theta_1} + \frac{\partial^2 \psi_2}{\partial^2 \theta_2} \right), \gamma_1 = \frac{1}{2} \left( \frac{\partial^2 \psi_1}{\partial^2 \theta_1} - \frac{\partial^2 \psi_2}{\partial^2 \theta_2} \right), \gamma_2 = \frac{\partial^2 \psi_1}{\partial \theta_1 \partial \theta_2} \quad (1.7)$$

Since both  $\kappa$  and  $\gamma_i$  are connected to the lensing potential, it is possible to reconstruct  $\kappa(\boldsymbol{\theta})$  from the  $\gamma_i(\boldsymbol{\theta})$  estimated by measuring galaxy shapes. In the flat sky approximation, the connection is straightforward in Fourier space, where the differential relations turn into algebraic ones, yielding

the Kaiser-Squires (KS) inversion [9]:

$$\tilde{\kappa}(\ell) = \frac{(\ell_1^2 - \ell_2^2) \tilde{\gamma}_1(\ell) + 2\ell_1\ell_2\tilde{\gamma}_2(\ell)}{\ell^2} \quad (1.8)$$

where  $\tilde{x}$  represents the Fourier transform of  $x$ , and  $\ell$  is the wave vector (large  $\ell$  corresponding to small angles on the sky). For surveys covering a large fraction of the sky, and complicated masks, more sophisticated algorithms are used [10, 11, 12, 13].

The convergence,  $\kappa(\theta)$  carries the same information as the shear field, and has a more direct link with physical models, since it can be defined as a weighted projected mass density:

$$\kappa(\theta) = \int d\chi W(\chi) \delta(\chi, \chi\theta) \quad (1.9)$$

where  $\chi$  is the comoving distance to the mass inhomogeneities that source the lensing potential,  $\delta$ , and  $W$  is a weight function that takes in to account the distribution along the line of sight of the galaxies whose images are being lensed,  $n_s(\chi_s)$ ,

$$W(\chi) = \frac{3\Omega_m H_0^2}{a(\chi)} \chi \int d\chi_s n_s(\chi_s) \frac{\chi_s - \chi}{\chi_s} \quad (1.10)$$

In this equation,  $\Omega_m$  is the present-day matter density in units of the critical density,  $H_0$  the Hubble constant and  $a$  the scale parameter.

Convergence maps,  $\kappa(\theta)$  have imprinted cosmological information from the evolution of the dimensionless matter over-density field,  $\delta$ , and the geometrical factors from the various distances involved. This Thesis deals with the use of  $\kappa(\theta)$  to infer the values of the cosmological parameters that determine the evolution of the universe, with a focus on non-Gaussian statistics measured at small angular scales. The methods discussed can also be applied to other cosmological probes.

### 1.1.2 Experimental results

From the first detection of shear induced by large scale structure [14, 15, 16, 17], surveys have been increasing in their sky coverage and depth, tightening the cosmological constraints derived

solely from WL measurements, and providing additional data for multi-probe analyses [18]. The analysis of the measurements has improved in parallel, especially in the area of systematics control. For a summary of experimental results up to 2014, see [5].

The state of the art in terms of sky coverage is the  $1321 \text{ deg}^2$  area utilized by the Dark Energy Survey Year 1 results (DES [19]), and in terms of survey depth, a galaxy surface density of  $\sim 17 \text{ arcmin}^{-2}$  was reached in the smaller  $137 \text{ deg}^2$  area used in the Subaru Hyper Suprime-Cam first year data (HSC [20]). Both benchmarks will soon be surpassed by upcoming releases from the same experiments, and future experiments at different stages of development. The Vera Rubin Observatory (VRO) Legacy Survey of Space and Time (LSST) imaging survey will cover  $18000 \text{ deg}^2$  with an effective survey depth of  $\sim 40 \text{ arcmin}^{-2}$  for its “gold” sample [21]. From space, the Euclid mission will reach an effective number density of galaxies of  $30 - 40 \text{ arcmin}^{-2}$  over  $\sim 20000 \text{ deg}^2$  [22]. Both experiments will benefit from each other (e.g. using high-resolution space-based data to improve de-blending of LSST images [23]), and the synergies would also apply to the Wide Field Infrared Survey Telescope (WFIRST [24]), which if approved, will reach an effective galaxy density of  $\approx 50 \text{ arcmin}^{-2}$  over  $2200 \text{ deg}^2$ .

The results from DES and HSC are consistent with each other and with previous WL experiments, such as the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS [25]) and the Kilo-Degree Survey (KiDS [26]), as shown by recent joint analyses [27], and may be in tension with measurements from *Planck* (albeit the tension still has a low statistical significance, see Fig. 1.1), making future measurements with higher constraining power even more relevant.

## 1.2 Cosmological inference based on WL measurements

The basic objective of cosmological inference is to estimate the posterior probability distribution of the parameters,  $\mathbf{p}$ , that define a candidate model,  $\mathcal{M}$ , to explain how the universe evolves, given some observations condensed in a data vector,  $\mathbf{d}$ .

$$P(\mathbf{p}|\mathbf{d}, \mathcal{M}) \tag{1.11}$$

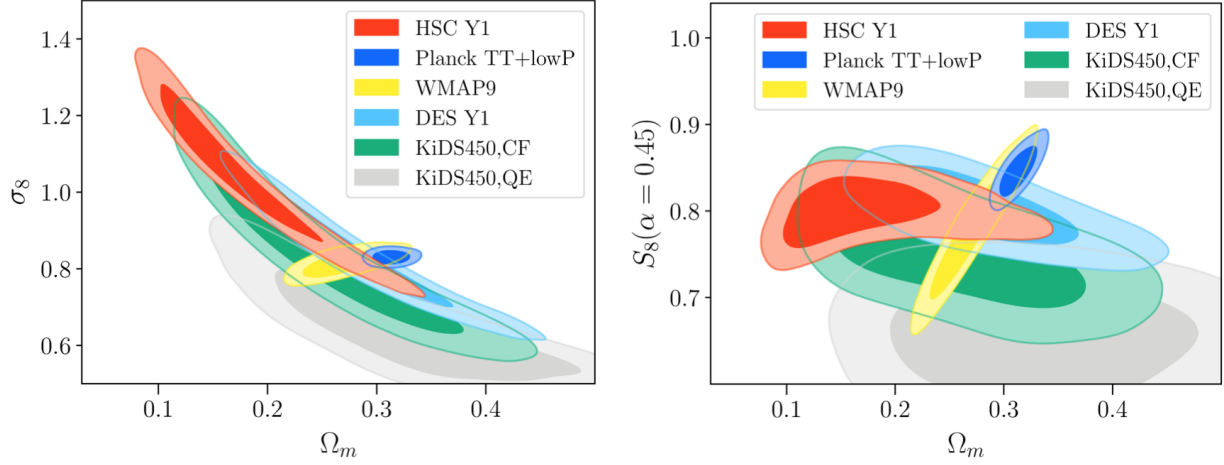


Figure 1.1: Comparison of parameter constraints from a series of WL analyses of galaxy surveys and *Planck*. (Figure from [20]).

Bayes theorem relates the probability distribution of the parameters given the observations, or posterior ( $P(\mathbf{p}|\mathbf{d}, \mathcal{M})$ ), to the more directly estimated probability distribution of the observations given the parameters, or likelihood ( $P(\mathbf{d}|\mathbf{p}, \mathcal{M})$ ):

$$P(\mathbf{p}|\mathbf{d}, \mathcal{M}) = \frac{P(\mathbf{d}|\mathbf{p}, \mathcal{M}) P(\mathbf{p}|\mathcal{M})}{P(\mathbf{d}|\mathcal{M})}. \quad (1.12)$$

The prior can be given by previous experiments. In the absence of such, an uninformative prior can be used instead, since in most cases the likelihood will dominate over the prior. In this Thesis, we restrict ourselves to comparing different data vectors to fully utilize WL observations. We can then ignore prior knowledge on the parameters,  $P(\mathbf{p}|\mathcal{M})$ , and the evidence for the data,  $P(\mathbf{d}|\mathcal{M})$  (we are not comparing models), and focus on the likelihood. The likelihood can be computed with a forward model that includes both the physical relationship between the parameters and the desired data vector, and the uncertainties of the measurement. For example, in many cases of interests –including those treated in the following chapters– the likelihood is a Gaussian on the data vector of dimension  $N_d$ :

$$P(\mathbf{d}|\mathbf{p}, \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^{N_d} |C|}} \exp\left(-\frac{1}{2} (\mathbf{d} - \langle \mathbf{d}(\mathbf{p}) \rangle)^T C^{-1} (\mathbf{d} - \langle \mathbf{d}(\mathbf{p}) \rangle)\right) \quad (1.13)$$

The data vector,  $\mathbf{d}$  can be either a statistic measured on simulated WL maps, or as we will see in Chapter 6, the output of a neural network whose parameters have been fit using simulated WL maps. In both cases, the physics of the forward model, and any observational and instrumental effects, are encoded in the simulated maps. The measurement of statistics (features) or use of a neural network provide a compressed representation of the data which, hopefully, preserves as much information as possible. The expected value of the data vector,  $\langle \mathbf{d}(\mathbf{p}) \rangle$ , the covariance matrix of the data vector,  $C$ , and the likelihood, are all estimated from ensembles of simulations.

### 1.2.1 WL statistics

The most commonly-used data vector to describe WL observations is the two point correlation function, which for the convergence field is defined as a spatial average over angular positions

$$\xi_{\kappa_i \kappa_j}(\theta) = \langle \kappa_i(\boldsymbol{\theta}_1) \kappa_j(\boldsymbol{\theta}_2) \rangle \quad (1.14)$$

with  $\theta = |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$ , and  $i, j$  two galaxy redshift bins. In the absence of a complex window function for the survey, the correlation function's Fourier transform, or power spectrum, is often a more convenient representation of the data

$$\langle \tilde{\kappa}_i(\boldsymbol{\ell}_1) \tilde{\kappa}_j^*(\boldsymbol{\ell}_2) \rangle \equiv \left( 2\pi^2 \right) \delta_D(\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2) P_{\kappa_i \kappa_j}(\ell) \quad (1.15)$$

with  $\ell = |\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2|$  and  $\delta_D$  the Dirac's delta function. Gaussian random fields are statistically fully determined by their power spectrum. Since the inhomogeneities in the early universe were Gaussian, and the early growth of structure, being linear, preserves their Gaussianity, it is natural to use the power spectrum (or two point correlation function) to characterize the convergence field.

Later stages of non-linear growth of structure induce non-Gaussianities in the  $\kappa$  field, and two-point statistics are no longer sufficient. Non-linear evolution is important at scales in which the variance of the dimensionless matter overdensities is large compared to unity. This condition sets a characteristic length scale, that grows with time, below which the growth of structures cannot be

described by linear theory, and non-Gaussianities are important. In the local universe, this scale is of order  $\sim 10$  Mpc ( $\sigma_8 \sim 1$ ), the size of galaxy clusters, which corresponds to an angular scale of order  $\sim 10$  arcmin at a redshift of  $z = 1.0$ , or a multipole of  $\ell \sim 1000$ . Future galaxy surveys with a high effective number density of galaxies will allow the reconstruction of mass maps with angular resolutions of order  $\sim 1$  arcmin.

Higher-order correlation functions provide complementary information, but they are degraded by non-Gaussian errors [28, 29, 30], which can be difficult to measure (for example, the covariance for the three-point correlation function is sourced by the six-point correlation function). Alternatively, morphological descriptors can extract non-Gaussian information, and it is possible to estimate them (and their errors) reliably from simulations. These descriptors, some discussed in this thesis, include lensing peaks, minima, and Minkowski functionals.

Lensing peaks are local maxima of the convergence field, and the typical statistic associated with them is their abundance as a function of their height [31, 32], although their profile, and abundance as a function of their profile, have also been considered [33, 34]. Local minima can be used in a similar way as local maxima. They are especially interesting because they probe cosmic voids which are sensitive to DE, alternative gravitational models, and the masses of neutrinos [35, 36, 37]. Besides, they may be less sensitive than lensing peaks to systematic effects induced by baryonic physics (e.g. galaxy feedback) [38, 39]. Three Minkowski functionals (MFs) can be defined over excursion sets on a two-dimensional field. They correspond to the area, boundary length and boundary genus of the set of points whose value exceeds a given threshold. As for peaks, theoretical studies have shown that complementary information to that from second-order statistics can be accessed through the use of MFs [40, 41, 42]. MFs are expected to play a valuable role in future surveys [43] and in the exploration of new physics [44].

Non-Gaussian statistics have started to be used successfully to improve constraints derived from recent WL experiments. The CFHTLenS survey, with an effective galaxy number density of  $\approx 11 \text{ arcmin}^{-2}$  at a median redshift of 0.7, was the first WL experiment that offered a large field of view ( $149 \text{ deg}^2$ ), and has been analyzed extensively with non-Gaussian statistics. The

three-point correlation function improved by  $\approx 10\%$  parameter constraints derived from the two-point correlation function alone [45]. Lensing peaks have been measured in at least three different studies. First, an analysis of peak abundance and tangential shear profile around peaks, restricted to stripe 82 data ( $\approx 124 \text{ deg}^2$ ) could discriminate between profiles of DM halos [46]—rejecting, for instance, the singular isothermal sphere model. Two peak abundance studies demonstrated how constraints on  $\Omega_m$  and  $\sigma_8$  can be improved by a factor of  $\approx 2$ , relative to what can be achieved with the two-point correlation function or the power spectrum [47, 48]. Cosmological constraints from Minkowski functionals can be degraded by a factor of a few by observational biases [49], and be biased themselves [50]. Higher-order moments of the convergence offer an un-biased alternative to the Minkowski functionals [50].

There are fewer examples yet of the application of non-Gaussian statistics in the most recent experiments. The KIDS survey offered a larger footprint compared to CFHTLenS, and lensing peak analyses of data covering  $450 \text{ deg}^2$  yielded improvements of  $\approx 20\%$  over the parameter constraints derived from the two-point correlation function [51, 52]. Lensing peaks were also measured in the science verification data from the DES survey ( $129 \text{ deg}^2$ ), and showed in isolation similar constraining power as two-point statistics [53]. The DES first-year data release ( $1321 \text{ deg}^2$ ), has been used to measure the probability distribution function of the projected matter density field [54]. This statistic, which encodes the same information as the first Minkowski functional, yielded cosmological constraints that were compatible with those obtained from the analysis of galaxy and shear WL two-point functions. Finally, lensing peaks have been used to identify galaxy clusters on HSC data [55, 56], but not as a cosmological probe by themselves. Deep surveys such as HSC can be used to reconstruct high-resolution  $\kappa$  maps which would benefit greatly from the use of non-Gaussian statistics.

### 1.3 Deep learning applied to weak lensing cosmology

The interest about, use of, and research on neural networks (NNs), has exploded in the past few years. While models inspired in biological neurons were developed as early as in the 1940’s [57],

and arranged in networks to perform classification tasks in the 1950's [58], it was not until after 2012 that the field exploded. That year, AlexNet, a convolutional neural network (CNN), won the ImageNet Large Scale Visual Recognition Challenge [59]. Since then, neural networks with diverse architectures have shown super-human performance in a wide range of problems, from image classification to playing Go [60] and poker [61].

There are many extensive reviews of deep learning, a fast-evolving field [62, 63, 64, 65]. Here, we review some basics of deep learning models that are relevant for Chapters 6 and 7, and summarize how these models have been applied to the field of weak lensing cosmology.

### 1.3.1 Deep learning basics

#### *Neural networks*

Deep learning models are composed of multiple processing layers of connected nodes, or neurons, and can learn complex representations of data. Each neuron transforms its input, or pre-activation  $a$ , by a non-linear activation function,  $h$ :

$$z = h(a) \tag{1.16}$$

Typical activation functions include sigmoid-like functions that mimic the behavior of a system which exhibits a threshold sensitivity to inputs and output saturation, such as  $h(a) = 1/(1 + \exp(-a))$ ,  $h(a) = \tanh a$ ,  $h(a) = \arctan a$ ,  $h(a) = \text{erf}(a)$  or  $h(a) = H(a)$  (step function). Nowadays, the ‘‘Rectified Linear Unit’’ (ReLU),  $h(a) = \max\{0, a\}$  is commonly used.

Neurons are arranged in layers. They transform linear (or affine) combinations of the activations from the preceding layer and feed their outputs to the subsequent layer. For example, the output of a two-layer network (one hidden layer, 1, followed by the output layer, 2) can be computed as

$$y_k(\mathbf{x}, \mathbf{w}, \mathbf{b}) = h\left(\sum_j^M w_{jk}^2 h\left(\sum_i^N w_{ij}^1 x_i + b_j^1\right) + b_k^2\right) \tag{1.17}$$



where an input vector  $\mathbf{x}$  of dimension  $N$  is fed to  $M$  hidden neurons arranged in layer 1, defined by the weights  $w_{ij}^1$  and biases  $b_j^1$ . Layer 1 neurons transform their output with the activation function  $h$  and feed it to neurons in layer 2 (the output layer), which in turn transform linear combinations of the outputs of layer 1 neurons defined by weights  $w_{jk}^2$  and biases  $b_k^2$  according to the activation function  $h$ . In this simple structure, neurons in a given layer are connected to all the neurons in the preceding and subsequent layers (fully connected layers) but not to nodes within their own layer. The activation function does not need to be the same for all neurons in the network.

Neural networks can then be viewed as parametric, non-linear functions, whose calculation is equivalent to the forward processing of information (feed-forward network), from the input to the output layer, through a deterministic, directed acyclic graph. It is common to represent neural networks by these graphs, and use their graph representation to implement calculations (see, for example, the `TensorFlow` computing framework [66]).

The mapping of the input onto the output is defined by the weights and biases  $\mathbf{w}$  and  $\mathbf{b}$ . These parameters can be determined from examples of input data for which the correct output is known (supervised learning). A network with a single hidden layer, such as the one described by Eq. 1.17 can approximate any continuous function on a compact subset of  $\mathbb{R}^N$  [67]. Unfortunately, the required number of neurons to do so,  $M$ , can become exponentially large [68]. Multi-layer networks can be trained with available learning algorithms and offer the same ability to approximate functions with a bounded number of neurons [69]. Thus, all neural networks used in practice are multi-layered. When the number of layers exceeds a few, networks are referred to as deep neural networks (DNNs).

While the precise shape of the activation functions used does not limit the ability of a network to represent functions, the functions must be non-linear. Otherwise, a linear combinations of linear functions is itself a linear function, and the network reduces to a linear classifier.

### Network training

Given a set of input examples with known outputs (training data set), the parameters that characterize the network can be determined by minimizing a loss function defined over the training data set. A common loss function is the average of the quadratic distance between the network's output and the true (known) mappings for the examples in the training data set (mean square error loss, or MSE). If we combine the weights and biases of the network in a single parameter vector,  $\mathbf{p}$ , this loss can be expressed as

$$L(\mathbf{p}; \{\mathbf{x}_i\}, \{\mathbf{t}_i\}) = \frac{1}{N} \sum_i^N \|\mathbf{y}(\mathbf{x}_i, \mathbf{p}) - \mathbf{t}_i\|^2 \quad (1.18)$$

Cast in this form, network training is an optimization problem. When the activation functions are differentiable, gradient-descent (GD) algorithms can be used. The gradient of the loss function w.r.t. the network's parameters is used to update the parameters in each optimization step,  $s$ :

$$\mathbf{p}^{s+1} = \mathbf{p}^s - \eta \nabla_{\mathbf{p}} L(\mathbf{p}^s) \quad (1.19)$$

where  $\eta$  is the “learning rate”, and can be adjusted during the optimization process. The gradient can be evaluated using the whole training data set, or just a sub-set of it (stochastic gradient descent, or SGD).

For a feed-forward neural network, the gradient of the loss function can be efficiently computed using the error backpropagation algorithm [70]. SGD can help avoid local minima of the non-convex loss function. An activation function that accelerates the learning process is the ReLU, a piece-wise linear function which prunes the negative part to zero leaving the positive part untouched [71]. A slight modification, with a small, non-zero slope in the negative domain (Leaky ReLU [72]) prevents neurons from getting stuck with zero gradients.

### *Learning symmetries: convolutional neural networks*

In many cases, the desired output of a NN must be invariant to transformations of the input data. For instance, cosmological parameters inferred from convergence maps should be invariant to arbitrary rotations on the sphere (or translations and plane rotations in the flat sky approximation). In general, there are four ways to achieve the NN invariance w.r.t. transformations of the input data:

- Train on features that are invariant to the transformation. For example, the  $\kappa$  power spectrum of a map is invariant to the symmetries of the map, so a network trained on power spectra will exhibit the same invariance. Any information lost in the feature extraction step will not be learned by the network.
- Provide a sufficient number of examples in training to allow the network to learn the underlying symmetries in the data. This can be achieved, for example, by augmenting the training data set, transforming each example according to a symmetry before feeding it to the network. For example, the same  $\kappa$  map can be used many times during training after applying a random rotation to enforce the rotation invariance of the NN.
- Modify the loss function with a term that penalizes changes in the NN output when the input changes following a given transformation. This technique, known as “tangent propagation” works for continuous transformations [73].
- Modify the structure of the NN so that it is invariant to specific transformations in the input. This usually results in additional conditions to be satisfied by the NN’s parameters, which reduce the model’s complexity.

The last approach is followed by a class of NNs which has been very successful classifying natural images: convolutional neural networks (CNNs [74]). CNNs modify the basic structure of fully connected networks in three ways. First, each neuron is not connected to all the nodes of the previous layer, but just to a small sub-sample of neighboring neurons. This local support

takes advantage of the strong spatial correlations between nearby pixels typical of natural images. Second, the learned weights are shared among all the neurons within a given layer. Weight sharing on a local support is equivalent to a convolution with a filter defined by the learned weights of the layer, hence the name of convolutional networks. The structure is invariant to translations. If a feature is useful for classification, it will probably be useful irrespective of its precise location in the image. An additional advantage is that it reduces the model complexity by shrinking the number of learned parameters. Finally, convolutional layers are mixed with sub-sampling steps. Sub-sampling allows the CNN to learn features on multiple scales without increasing the size of the convolutions used, as well as abstract high-level relationships from low-level features. The reduction in the dimensions of the feature maps as layers approach the output further lowers the model complexity, making training easier and providing some regularization.

The models used in Chapters 6 and 7 of this Thesis correspond to this class of networks. Since we apply them to simulated data in the flat sky approximation, we do not need to modify structures that are common in the field of image analysis. It is possible to modify these structures to account for a spherical geometry in the data. Such schemes can be useful to analyze surveys covering a large fraction of the sky. They often are invariant to rotations [75, 76], unlike standard convolutional layers. In Chapter 6, we augmented the training data to enforce rotation invariance of the CNN output to rotations on the plane. An alternative path would be to use Group Equivariant CNNs, which generalize the definition of convolution to take into account additional symmetries [77].

### *Challenges to the adoption of deep learning models to WL cosmology*

Two of the main barriers to the adoption of DNNs in the analysis of WL data result from their complexity: the risk of over-fitting and the interpretation of their output.

As the number of learned parameters in DNNs increases –some models are defined by millions of parameters– so do over-fitting concerns. Over-fitting in this context means that errors measured in the training data set underestimate errors on data not used during training. That is, the model

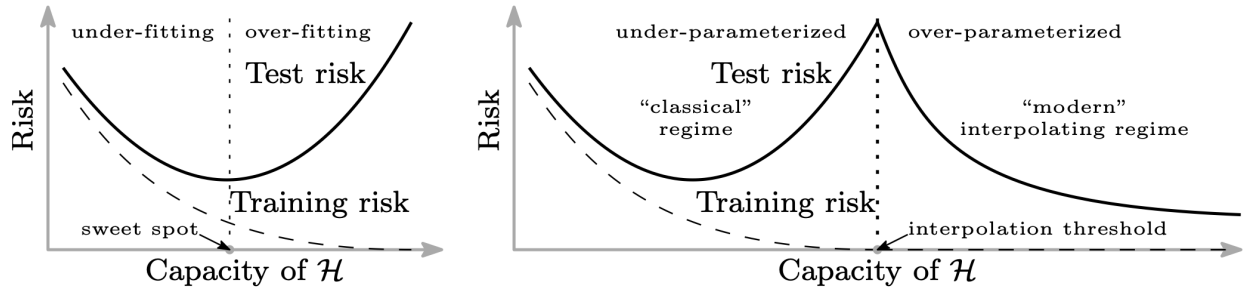


Figure 1.2: Comparison of bias-variance trade-off for traditional machine learning algorithms on the left, and the different regime for DNNs on the right (Figure from [78]).

does not generalize well to new data. Over-fitting is related to the known bias-variance trade-off in machine learning: as a models' complexity increases, it fits better the training data set at the cost of worsening its ability to generalize.

An example that illustrates this phenomenon is the interpolation of a set of points  $\{x_i, y_i\}$  using polynomial functions. Any training data set can be fitted with zero error using a high-enough degree for the polynomial. To achieve that, the polynomial's coefficients typically adopt large values, resulting in high-amplitude oscillations in the domain of  $x$  not constrained by the training data set. Therefore, new data points will be fit poorly. There is a sweet spot in terms of complexity: large enough to express the overall structure of the data but small enough to avoid unjustified large oscillations (see left panel of Figure 1.2).

DNNs seem to violate this trade-off, as they achieve essentially zero training error without sacrificing their ability to generalize (see right panel of Figure 1.2). Numerical experiments suggest that the extremely large number of parameters that define DNNs allow them to interpolate training data sets while keeping the norm of  $\mathbf{p}$  bounded [78], so long as the DNNs are fit using GD or SGD. The small norm of the parameters prevent the large oscillations in the output responsible for poor fitting of data not present in the training data set. The large number of parameters effectively regularizes DNNs.

The second barrier to the adoption of DNNs is that, as the number of layers increases, non-linear models become hard to interpret. Understanding what features in the data drive the output of DNNs would increase the trust in DNNs, especially since they are capable of learning random

labels [79]. Furthermore, for applications in cosmology, mapping the models' outputs to features would allow to understand the impact of specific physical structures on the lensing signal, assess the robustness of models' predictions, and engineer alternative, less computationally expensive, methods (e.g. summary statistics).

For a detailed review on explainable artificial intelligence (AI) see [80]. In general, the methods developed fall into one of the following categories:

- Use of surrogate models. These methods evaluate DNNs around an input of interest, and approximate it with another model (e.g. linear classifier) for which the interpretation is straightforward. Examples of this approximations include the technique known as Local Interpretable Model-agnostic Explanations (LIME [81]).
- Local input perturbations. By analyzing the changes in the output when specific features of the input are perturbed (e.g. masked), it is possible to assess the relevance of these features on the model [82, 83].
- Feature visualization. Inspecting which inputs maximize the activation of specific neurons of layers within a model can help assess its relevance [84].
- Sensitivity methods try to explain how the output of a DNN changes as the input features are perturbed. They are generally based on the gradients of a network's output w.r.t its input [85].
- Propagation-based methods use the internal structure of a model to distribute its output in input space. Examples of these methods include Layer-wise Relevance Propagation (LRP [86]), Deconvolution [87], and Guided Backpropagation [88].

Some of these methods have been shown to yield interpretations that are insensitive to either the input vector or the model's parameters, raising questions regarding their validity [89]. It is critical to test these methods for the combination of data and model of interest before using them to evaluate which input features are more relevant for the DNN output. Chapter 7 shows an application of these methods to interpret a DNN trained on simulated WL data.

### 1.3.2 Applications of deep learning to WL

The transition of cosmology from a data-deprived to a data-rich discipline has resulted in the increasingly important role played by data analysis techniques. This is not only true for deep learning algorithms, but for machine learning in general [90, 91]. Within the restricted field of WL cosmology, deep learning has already been used in almost every step of the chain that links the acquisition of images in galaxy surveys to the inferred cosmological models that best fit the data.

CNNs have been trained to learn the point spread function (PSF) of the instruments used to map the shapes of galaxies [92]. The measured shape of galaxies is the convolution of their true apparent shape and the PSF. Galaxy shapes need to be parametrized from a set of pixel intensities. This task has also been performed using CNNs [93], improving computation times by four orders of magnitude. The physical magnitude of interest, the lensing shear, has been estimated without bias from neural networks [94], by training them on combinations of the galaxies' light intensity moments [95], or the raw images themselves [96].

Another area where DNNs have been used extensively is the estimation of the redshift of galaxies given only color filter imaging [97, 98, 99, 100, 101, 102, 103, 104, 105], with the aim of minimizing one of the main sources of systematic errors in photometric galaxy surveys.

Galaxy shapes and photometric redshifts can be combined to build mass maps, a step that can be performed with convolutional networks [106]. Generative Adversarial Neural Networks (GANs [107]), a model that combines two neural networks (one as a classifier and the other as a generator), trained until they reach a Nash equilibrium in which the classifier cannot tell whether an input has been created by the generator, can be used to de-noise those reconstructed mass maps [108, 109], or as computationally cheap, fast emulators of synthetic maps. GAN-based emulators offer high-quality additional realizations from cosmologies present in the training data sets [110, 111], and have also the ability to generalize to new cosmologies [112].

Lastly, DNNs can be used to discriminate between cosmological models from measured/simulated mass maps (cosmological inference). CNNs have been shown to break the  $\Omega_m - \sigma_8$  degeneracy in WL maps [113], provide parameter constraints competitive with those obtained from alternative

methods in noiseless data [114, 34], noisy data [115, 116], and survey data [117]. They have also been proven useful to discriminate modified gravity cosmologies [118, 119].

## 1.4 Structure of dissertation

This thesis examines different aspects of the challenges to extract and use small-scale information from near-future weak lensing data sets. The remaining chapters are organized as follows:

Chapter 2 disentangles the relative contribution of the expansion history of the universe (geometry) and the growth of cosmic structures (growth) to the sensitivity of non-Gaussian statistics to cosmology. The first alters the relative distances between sources, lenses, and observers. The second affects the evolution of the lensing potentials. Both impact the overall deflection of photons propagating through the inhomogeneous universe. This exercise is accomplished by ray-tracing “mixed” past light cones, composed of lensing planes at distances that correspond to a different cosmology to the one used to evolve the lensing potentials. The method is tested with the power spectrum, for which the split can be computed analytically. When applied to non-Gaussian statistics –the equilateral bispectrum, lensing peaks, and Minkowski functionals– it shows how both effects tend to cancel each other. This cancellation affects distinct statistics and cosmological parameters differently. The equilateral bispectrum is the worst affected observable, and the sensitivity of non-Gaussian statistics to the equation of state of DE ( $w$ ) is reduced twice as much as that to the density of the universe ( $\Omega_m$ ). This chapter has been published as [120].

In Chapter 3, we assess the performance of a semi-analytic, halo-based model for the counts of lensing peaks: CAMELUS. In principle, such a model is an attractive alternative to computationally expensive simulations of non-linear growth of structure, especially since the sensitivity of lensing peaks to some cosmological parameters ( $\Omega_m$ ) is dominated by growth (see Chapter 2). When compared with measurements over ray-traced simulations, CAMELUS fails to account for correlations in the halo positions, resulting in somewhat under-estimated covariance matrices that are less sensitive to cosmological parameters. Mean peak counts from both methods show disagreements as well, notably in the negative tail and low significance regime. While the addition of shape



noise mitigates these disparities, further improvements are needed to use safely this semi-analytic method for inference. This chapter has been published as [121].

Relying on numerical simulations for the prediction of non-Gaussian statistics and their covariances, highlights the importance of reducing their computational cost. Chapter 4 uses a series of numerical experiments to define some minimum requirements of these simulations for a LSST-like survey in terms of sky coverage and tomographic galaxy density. A mass resolution of  $7.2 \times 10^{11} h^{-1} M_{\odot}$  can be used to forecast measurements of the PDF, lensing peaks, and Minkowski functionals. This is an order-of-magnitude lower than resolutions typically used for theoretical studies. Another hyper-parameter that significantly impacts the cost of simulations based on the multi-lens-plane algorithm, is the thickness of the lensing planes used to build past light cones. While reducing it allows to recycle simulations into many pseudo-independent light cones, the loss of power induced by the lensing plane’s window function propagates into non-Gaussian observables. In practice, this sets a minimum lensing plane thickness of  $\approx 60 h^{-1} \text{Mpc}$  for the considered survey specification. This chapter has been published as [122].

Chapter 5 evaluates the detection feasibility of the kinematic Sunyaev-Zeldovich effect induced by the rotation of gaseous galactic halos. The ability to model the impact of baryonic processes on the distribution of matter at small scales ( $\approx 1 \text{ Mpc}$ ) limits the amount of additional cosmological information that can be gained from non-Gaussian statistics. We study the potential of combining future high-resolution CMB data with galaxy spin measurements from integral field spectroscopic surveys to statistically detect the CMB temperature dipole induced by the rotation of ionized galactic halos. In the near future, this novel probe can be used in combination with other observations to constrain models of the gas distribution around galaxies, improving cosmological inference from weak lensing, and informing models of galaxy evolution. This chapter has been published as [123].

Weak lensing summary statistics compress the information present in the original data (shear catalog or  $\kappa$  map). In general, this compression is not lossless. An alternative path is to train a model to learn the non-linear mapping between the original data and the cosmological parameters of interests. In Chapter 6 we apply deep learning for the first time to pixel-level simulated weak

lensing data, and show that a trained network can extract more cosmological information than lensing peaks, a competitive non-Gaussian statistic. Deep learning can be used to estimate the information content of future data sets, as a benchmark for more traditional inference methods, or as an independent estimator. This chapter has been published as [\[114\]](#).

Chapter 7 tries to interpret the outputs from deep neural networks applied to weak lensing data. While DNNs seem to access information in excess of what can be reached with non-Gaussian statistics, it is important to understand which features in the data drive the networks' output. This will prevent inference based on spurious elements, assess the robustness of the inference to systematic effects, and inspire the design of more efficient non-Gaussian statistics.

Finally, Chapter 8 summarizes the key results presented in this Thesis and outlines possible directions for future work.

## Chapter 2: Geometry and growth contributions to cosmic shear observables

### 2.1 Introduction

A cosmological model with a nearly scale-invariant primordial fluctuation spectrum, cold dark matter (CDM) and dark energy (DE) matches well a wide range of observations, from the Universe’s expansion measured by standard candles [1, 2] and standard rulers [124], to its primordial chemical composition [125, 126], structure formation and the properties of the Cosmic Microwave Background (CMB) [127]. While non-baryonic DM and DE make up most of the present-day energy density of the Universe [128], the nature of either dark component remains unclear.

Cosmic shear is the weak gravitational lensing of background sources by large scale structure [4, 5]. It probes the matter density field through the gravitational potential fluctuations, and is also sensitive to the expansion history of the Universe through the distances between the observer, lensed source and lensing structures. While lensing is usually characterized by a measurement of the shear through the shapes of background galaxies, convergence (magnification) statistics can be inferred from these measurements, and are considered here for ease of computation. The polyspectra of the convergence field are equal to the E-modes of the shear field.

Ongoing and upcoming surveys, such as the Dark Energy Survey (DES <sup>1</sup>), the Large Synoptic Survey Telescope (LSST <sup>2</sup>), the Euclid mission <sup>3</sup> and the Wide Field Infrared Survey Telescope (WFIRST <sup>4</sup>), include weak lensing in their scientific program as part of their effort to test the concordance model with unprecedented precision and shed light on the nature of DM and DE. To realize this potential, we need observables that extract all the cosmological information from the data, as well as models capable of predicting them with high accuracy.

---

<sup>1</sup><http://www.darkenergysurvey.org>

<sup>2</sup><http://www.lsst.org>

<sup>3</sup><http://sci.esa.int/Euclid/>

<sup>4</sup>[wfirst.gsfc.nasa.gov](http://wfirst.gsfc.nasa.gov)

Second-order statistics do not fully capture non-Gaussianities in the lensing signal from non-linear gravitational collapse on small scales. Numerous alternative observables have been proposed to extract this extra information, from higher-order correlation functions [129, 130] and moments [131] to topological features like local maxima (peaks) [132] and Minkowski functionals [133].

In this work, our goal is to clarify the sensitivity of such observables to the expansion history of the Universe (“geometry”) and to the evolution of primordial inhomogeneities into cosmic structures (“growth”). The analogous question has been addressed for the convergence ( $\kappa$ ) power spectrum [134]. The geometry vs. growth decomposition of the power spectrum has improved our understanding of constraints on DE from weak lensing [135], provided an alternative cosmological probe independent of the growth of structures [136, 137], has been used to strongly constrain deviations from general relativity [138] and has allowed a consistency test of the standard cosmological model [139].

Our work extends previous studies to observables beyond the power spectrum. In particular, we analyze the equilateral bispectrum and two simple but promising topological observables: lensing peaks and Minkowski functionals. We restrict our analysis to two parameters that can influence lensing observables significantly through both geometry and growth: the total matter density ( $\Omega_m$ ) and the DE equation of state as parametrized with a constant ratio of its pressure to its energy density ( $w$ ). Future work should include a full cosmological parameter set. We disentangle the two contributions by measuring observables over a collection of mock  $\kappa$  maps built from ray-tracing N-body simulations.

The paper is organized as follows. In § 2.2, we describe the suite of simulations we used and our method to separate the effects of geometry and growth on the observables. In § 2.3, we show the sensitivity of each observable to both  $\Omega_m$  and  $w$ , discussing the separate contributions from geometry and growth, and in § 2.4 we show how they impact parameter inference. We then discuss our results in § 2.5 and summarize our conclusions in § 2.6.

$\Omega_m$	$w$	$\Omega_m$	$w$
0.20	-1.0	0.26	-0.5
0.23	-1.0	0.26	-0.8
0.29	-1.0	0.26	-1.2
0.32	-1.0	0.26	-1.5

Table 2.1: *Parameters of the eight models explored around the fiducial model ( $\Omega_m = 0.26$ ,  $w = -1.0$ ). All models are spatially flat with  $\Omega_\Lambda = 1 - \Omega_m$  and consider a constant equation of state parameter  $w$  for DE.*

## 2.2 Disentangling geometry from growth in simulations

We measured lensing observables on mock  $\kappa$  maps generated for 9 flat  $\Lambda$ CDM cosmologies. We considered only DE models with a constant ratio of pressure to energy density ( $w$ ). Apart from  $w$ , we also varied  $\Omega_m$ , with a fiducial model corresponding to  $\{\Omega_m, w\} = \{0.26, -1.0\}$  and the remaining 8 cosmologies each differing from it in just one parameter (see Table 2.1). For all models, we fixed the amplitude of perturbations at  $\sigma_8=0.8$ , the Hubble constant to  $h = 0.72$ , the spectral index to  $n_s=0.96$  and the effective number of relativistic degrees of freedom to  $N_{\text{eff}} = 3.04$ .

### 2.2.1 Simulating weak lensing maps

A set of mock convergence maps was generated by raytracing through the outputs of dark matter-only N-body simulations, following the multiple lens plane algorithm implemented in LENSTOOLS. We used full ray-tracing to avoid any potential bias in the convergence descriptors under study. While it has been shown that the Born approximation is accurate for the galaxy lensing power spectrum [140] and bispectrum [141], it can introduce significant biases for higher-order moments [142] and its effects on topological descriptors are yet unclear. We give a brief outline of our simulation pipeline here, and refer readers for a detailed description in [143].

The observer’s past light cone is discretized in a set of lens planes separated by a constant comoving distance of  $80 h^{-1}\text{Mpc}$ . For each cosmology, we evolved the matter density field in a single box of side  $240 h^{-1}\text{Mpc}$ , which can cover a field of view of  $3.5 \times 3.5 \text{ deg}^2$  up to a redshift  $z \approx 3$ . The N-body simulations were run using GADGET2 [144] with the same initial conditions.

Each box contains  $512^3$  particles, yielding a mass resolution of  $\approx 10^{10} M_\odot$ . All simulation volumes were randomly shifted and rotated to generate 1024 different  $\kappa$  maps for each cosmology. This is justified by previous work [145], which has shown that a single N-body simulation can be recycled to generate as many as  $\approx 10^4$  statistically independent realizations of the projected 2D convergence field.

Bundles of  $1024 \times 1024$  uniformly distributed rays were traced back to the lensed galaxies' redshift and the convergence was reconstructed from the accumulated deflection of the rays by the discrete lens planes. For simplicity, we assumed all source galaxies are uniformly distributed at a single redshifts, chosen to be either  $z_s = 1$  or  $z_s = 2$ .

We included the effect of galaxy shape noise assuming it is uncorrelated with the lensing signal and its probability distribution function (PDF) is a Gaussian with zero mean. The variance of the shape noise depends on the r.m.s. intrinsic ellipticity noise ( $\sigma_\epsilon$ ), the source galaxy surface density ( $n_{gal}$ ) and the pixel size ( $\theta_p$ ), as [147]

$$\sigma_p^2 = \frac{\sigma_\epsilon^2}{2n_{gal}\theta_p}. \quad (2.1)$$

For this work we considered an intrinsic ellipticity noise of  $\sigma_\epsilon = 0.4$  and a galaxy density of  $n_{gal} = 25 \text{ arcmin}^{-1}$ , similar to the expectation for LSST but conservative compared to the galaxy densities expected in deeper surveys, such as Euclid and WFIRST. We generated a single set of 1024 noise-only maps and added them to the noiseless  $\kappa$  maps ray-traced from the N-body simulations. We smoothed the noiseless  $\kappa$  and shape noise maps with the same 2D Gaussian kernel,

$$W(\theta) = \frac{1}{2\pi\theta_s^2} \exp\left[-\frac{\theta^2}{2\theta_s^2}\right], \quad (2.2)$$

with  $\theta$  the angular distance to each pixel, and a characteristic width  $\theta_s = 1 \text{ arcmin}$ . In this analysis we did not combine different smoothing scales. The smoothing suppresses power on small scales corresponding to spherical multipoles on the sky  $\ell \gtrsim 12000$ , which corresponds to the scale at

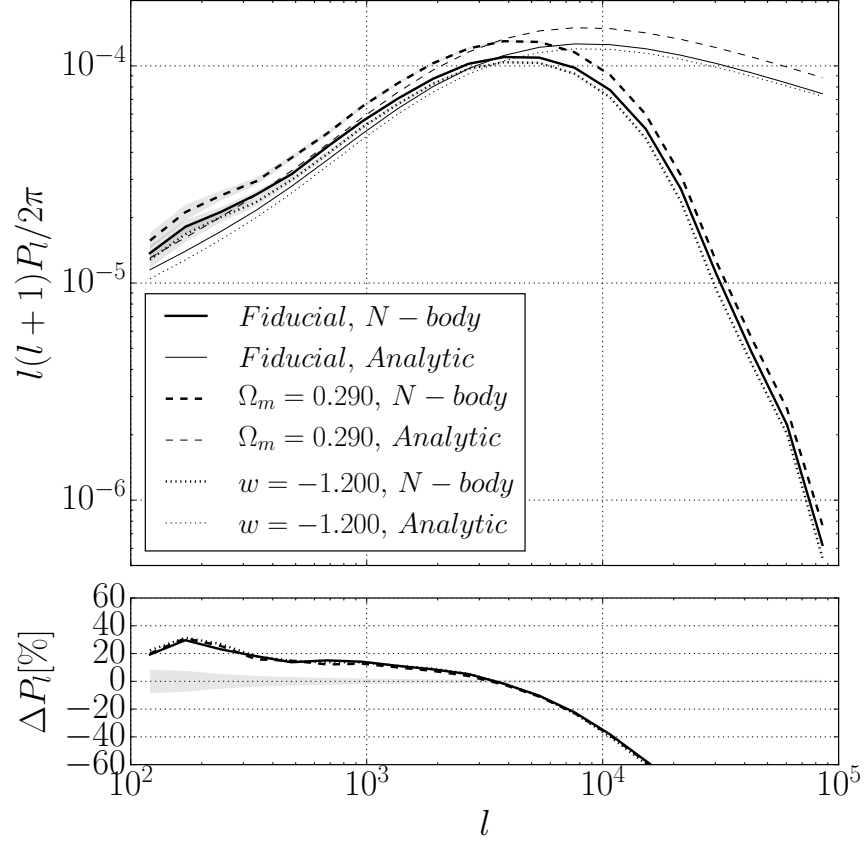


Figure 2.1: Comparison between the power spectra measured for selected models, as labeled, over noiseless, un-smoothed  $\kappa$  maps (thick lines) and analytic predictions using a fitting formula [146] for the matter power spectrum (thin lines). Percent differences between measured and predicted power spectra are depicted in the lower panel. Shaded areas represent  $\pm 1$  standard deviations around the average, scaled to a  $1000 \text{ deg}^2$  survey, and in the lower panel only the standard deviation for the fiducial model is plotted for reference.

which we are still not limited by the finite resolution of our simulations (see Fig. 2.1). We do not show results beyond  $l = 10000$ , and the topological features, measured on the smoothed maps, do not contain information from smaller scales.

### 2.2.2 Isolating the effect of geometry vs. growth

Galaxy shape distortions by gravitational lensing result from the convolution of the lens properties and the distances between source galaxies, lenses and the observer. Both effects depend on cosmology; the former through the evolution of mass inhomogeneities, and the latter through the expansion history of the Universe. To account for these effects separately in our simulations, we evolved the matter density field according to a cosmological model, but during the ray-tracing, we allowed distances to correspond to a different cosmology.

In our implementation of the multi-plane algorithm, lens planes are located at the same comoving distances from the observer for all models and we disentangled growth and geometry by modifying the lens planes' properties.

The lensing potential for a lens at a comoving distance of  $\chi_i$ , given a set of cosmological parameters  $\mathbf{p}$ , is determined by its mass surface density,

$$\Sigma_i(x, y; \mathbf{p}) = \frac{3H_0^2\Delta}{2c^2} \frac{\chi_i}{a(\chi_i, \mathbf{p})} \delta\Omega_m(x, y; z(\chi_i, \mathbf{p}); \mathbf{p}), \quad (2.3)$$

where  $(x, y)$  are angular positions on the lens plane,  $\Delta$  is the plane's thickness ( $80 h^{-1}\text{Mpc}$ ),  $\chi$  the comoving distance,  $a$  the scale factor and  $\delta\Omega_m$  the product of the density contrast and the matter density parameter. The sensitivity of an observable to cosmology refers to the change in that observable for a set of parameters  $\mathbf{p}$  relative to the same observable for a fiducial model  $\mathbf{p}_0$ .

The effect of geometry can be estimated by evolving the perturbations according to  $\mathbf{p}_0$  and evaluating them at redshift  $z(\chi_i, \mathbf{p})$ , keeping the geometrical prefactor  $\chi/a$  equal to the value that corresponds to the cosmological model  $\mathbf{p}$ . Conversely, the effect of the growth of structures can be captured by keeping the geometrical prefactor equal to its value in the fiducial model and evaluating



the density perturbations at  $z(\chi_i, \mathbf{p}_0)$  after evolving them according to  $\mathbf{p}$ .

$$\Sigma_i^{Geometry}(x, y; \mathbf{p}; \mathbf{p}_0) = \frac{3H_0^2\Delta}{2c^2} \frac{\chi_i}{a(\chi_i, \mathbf{p})} \delta\Omega_m(x, y; z(\chi_i, \mathbf{p}); \mathbf{p}_0) \quad (2.4)$$

$$\Sigma_i^{Growth}(x, y; \mathbf{p}; \mathbf{p}_0) = \frac{3H_0^2\Delta}{2c^2} \frac{\chi_i}{a(\chi_i, \mathbf{p}_0)} \delta\Omega_m(x, y; z(\chi_i, \mathbf{p}_0); \mathbf{p}) \quad (2.5)$$

This approach does not require running separate N-body simulations to generate growth-only and geometry-only convergence maps, but it involves saving additional GADGET2 snapshots, since fixed comoving distances correspond to different scale factors for different cosmologies. For each model  $\mathbf{p}$ , additional snapshots at redshifts  $z(\chi_i, \mathbf{p}_0)$  are needed. For the fiducial cosmology, we saved additional snapshots at redshifts  $z_k(\chi_i, \mathbf{p}_k)$  for each  $\mathbf{p}_k$  model considered.

### 2.3 Sensitivity to $\Omega_m$ and $w$

The percentage deviation of an observable relative to its value in the fiducial model measures its sensitivity to changes in cosmology. For galaxy lensing, we are interested in observables measured over  $\kappa$  maps that include shape noise. We focus on the behavior of four observables: the power spectrum, which has already been studied analytically and will serve as a test of our simulation-based approach, the equilateral bispectrum, which should be zero for a Gaussian random field, and two topological features that have been used to probe non-Gaussianities: lensing peaks and Minkowski functionals. We measured the sensitivities from the full ray-traced N-body simulations, as well as from simulations that only capture the changes due to either the expansion history or to the structure growth in a given cosmology.

### 2.3.1 Power spectrum

The convergence power spectrum is the Fourier transform of the 2-point correlation function of  $\kappa(x, y)$  and is one of the most popular weak lensing observables. For a flat cosmology, with lensed sources at a fixed redshift, and using the Limber and flat-sky approximations, the power spectrum can be expressed as a line-of-sight integral of the matter power spectrum, weighted by a geometrical kernel [148]

$$P_k(l) = \frac{9}{4} \left( \frac{H_0}{c} \right)^4 \Omega_m^2 \int_0^{\chi_s} \frac{d\chi}{a^2(\chi)} \left( 1 - \frac{\chi}{\chi_s} \right)^2 P_\delta \left( \frac{l}{\chi}; \chi \right) \quad (2.6)$$

Where  $\chi$  is the comoving distance and  $\chi_s$  the comoving distance to the lensed galaxies. Geometry affects the power spectrum through  $\chi$  and the scale factor  $a$ . Growth enters the above expression through the matter power spectrum,  $P_\delta$  (including non-linear effects), and the  $\Omega_m^2$  outside of the integral. For our analytic calculations, we used the NICAIA implementation of the convergence power spectrum with the prescription from [146] for the matter power spectrum.

We determined the percentage deviation of the power spectrum relative to the fiducial cosmology over 1024 noiseless, un-smoothed  $\kappa$  maps for each non-fiducial cosmology, and compared the results with analytic predictions. These results, shown in the upper panels of Fig. 2.2, match the analytic predictions within the statistical uncertainties, and are also in good agreement with the findings of [134]. The sensitivity is only weakly dependent on the multipole.

The sensitivity to  $\Omega_m$  is dominated by growth, with a  $\approx 25\%$  change that is what would be expected from the  $\approx 12\%$  change in  $\Omega_m$  ( $\Omega_m^2$  outside the integral in Eq. 2.6). Geometry acts in the opposite direction, reducing the overall sensitivity by  $\approx 20\%$ . The sensitivity to  $w$  is dominated by geometry. While we expected its sensitivity to be smaller than that to  $\Omega_m$  due to the integrating effect, the partial cancellation between growth and geometry is even more severe. It reduces the sensitivity further ( $\approx 50\%$ ) to a level of  $\approx 5\%$  for a  $20\%$  change in the parameter. The smaller sensitivity should propagate into tighter constraints on  $\Omega_m$  than on  $w$  from weak lensing data.

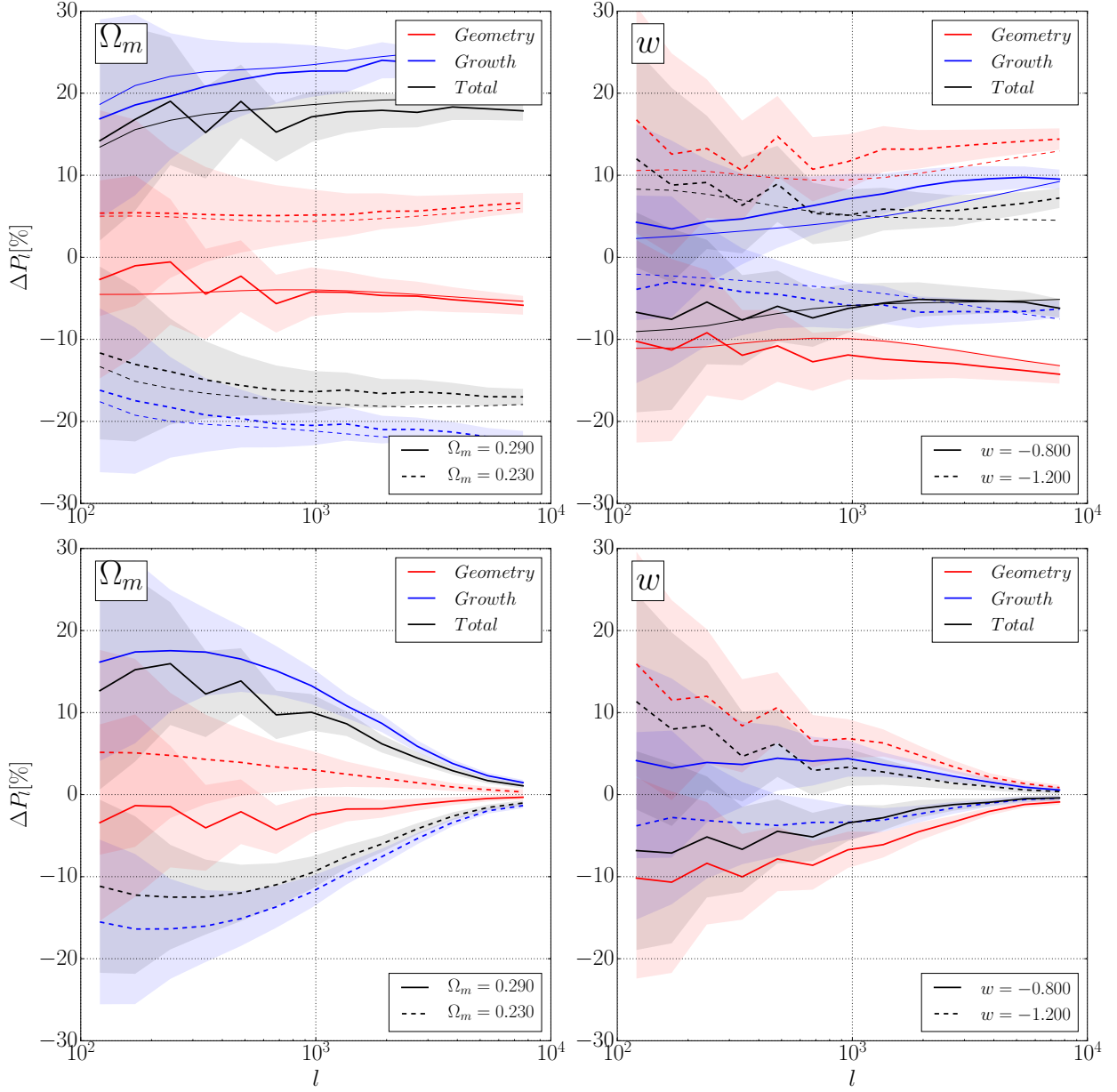


Figure 2.2: (color online) Sensitivity of the power spectrum to  $\Omega_m$  and  $w$  for noiseless (upper panels) and noisy (lower panels) convergence. Estimates including only geometry effects are shown in red, those including only growth effects in blue, and those including both effects in black. In the upper panels, analytic predictions are displayed with thin lines, for comparison. Source galaxies are at  $z_s = 1.0$  in all cases. Shaded areas represent a  $\pm 1$  standard deviation around the measured averages scaled to a survey sky coverage of  $1000 \text{ deg}^2$  and only selected models are displayed for clarity.

The origin of the partial cancellation is explained in detail in [134], but we reproduce the argument here for convenience. Making  $w$  more negative, from the fiducial  $w = -1.0$  to  $-1.2$ , yields a higher DE density in the past. The comoving distance to the source galaxies' redshift becomes larger, and so does the cumulative effect of small deflections experienced by light rays. As a result, the effect due to geometry is an increase of the lensing signal. Since we fix the amplitude of the perturbations at the present time ( $\sigma_8$ ) in our simulations, a higher DE density in the past means there are fewer structures to deflect the light rays in the past, and the growth contribution to the lensing signal is smaller compared to a model with constant dark matter density.

Galaxy shape noise introduces a scale-dependence to the relative sensitivity, as clearly seen in the lower panels of Fig. 2.2. At small scales, white noise dominates the power spectrum and suppresses its sensitivity to cosmological parameters. Galaxy shape noise then limits the information that can be extracted from the convergence power spectrum at small scales.

### 2.3.2 Equilateral bispectrum

The natural extension to the two-point correlation function is the three-point correlation function, or its Fourier transform, the bispectrum. A non-zero bispectrum is a clear non-Gaussian signal and has been detected in shear data [149, 45]. The analog of Eq. 2.6 links the convergence bispectrum to the bispectrum of the underlying matter density field through a Limber integration [148]

$$\mathbf{B}_k(l_1, l_2, l_3) = \frac{27}{8} \left( \frac{H_0}{c} \right)^6 \Omega_m^3 \int_0^{\chi_s} \frac{d\chi}{(\chi a(\chi))^3} \left( 1 - \frac{\chi}{\chi_s} \right)^3 \delta^D(l_1 + l_2 + l_3) \mathbf{B}_\delta \left( \frac{l_1}{\chi}, \frac{l_2}{\chi}, \frac{l_3}{\chi}; \chi \right) \quad (2.7)$$

Where  $\delta^D$  is a Dirac delta. When the lengths of the triangle defined by the three points on which the correlation function are measured are the same, the result is the equilateral bispectrum ( $\mathbf{B}_{lll}$ ). In an exercise analogous to the one done for the power spectrum, we measured  $\mathbf{B}_{lll}$  for our mock noiseless convergence maps and show their relative sensitivity to the cosmological parameters in Fig. 2.3.

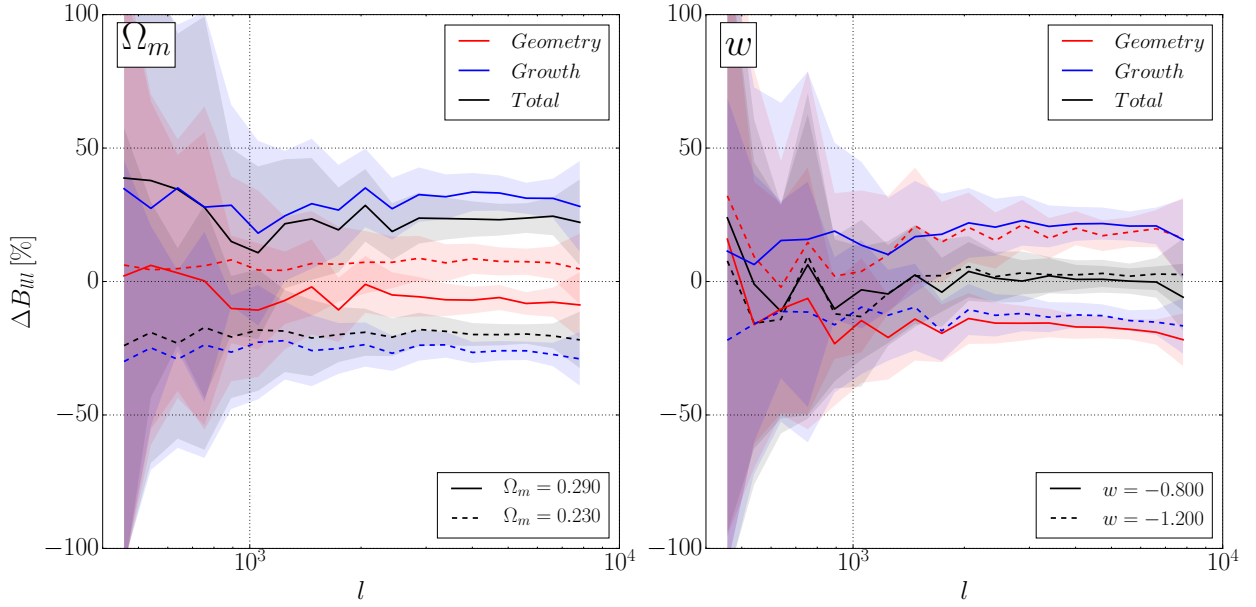


Figure 2.3: (color online) Sensitivity of the equilateral bispectrum of the noiseless convergence field to  $\Omega_m$  and  $w$ . Both panels show the percentage deviation in each model from the fiducial bispectrum. For clarity, only two models are depicted per panel, with the source galaxies at  $z_s = 1$ . As in Fig. 2.2, black lines show the net sensitivity, red lines the sensitivity due only to differences in geometry and blue lines the sensitivity due only to differences in growth. Shaded areas represent  $\pm 1$  standard deviation around the measured averages, scaled to a  $1000 \text{ deg}^2$  survey.

While noisier, the parameter-sensitivity has a behavior very similar to the case of the power spectrum, in terms of its weak dependence on the angular scale  $\ell$ , order of magnitude, and split between geometry and growth. The most noticeable difference is that the cancellation between both effects is almost perfect for  $w$ , resulting in a statistic that is almost insensitive to that parameter. The results for the lensed galaxies at  $z_s = 2$  are similar, and show the same cancellation for  $w$ . The addition of shape noise results in an even noisier measurement (see § 2.4) with error bars 3-4 times larger than the ones displayed in Fig. 2.3 for the noiseless case. There is no average sensitivity suppression at small scales, because the shape noise is Gaussian.

### 2.3.3 Lensing peaks

Peaks, defined as local maxima on smoothed  $\kappa$  maps, probe high-density regions, where non-Gaussianities of the convergence should be enhanced. Also, they are computationally inexpensive to measure, making them an attractive observable to combine with others for cosmological inference. Indeed, their distribution as a function of their height, or peak function, has been forecast to improve constraints obtained using only second-order statistics by a factor of 2 – 3 [32, 31]. Similar improvements have now been found in recent lensing survey data [47, 48, 53].

We extracted peak catalogues from our mock convergence maps and computed the percentage deviation of the peak height function relative to the fiducial model. The results for the noisy case are shown in Fig. 2.4. We again observe some similarities between the sensitivity of the peak height functions and that of the power spectrum. The  $\Omega_m$ -sensitivity is dominated by growth, while geometry dominates the sensitivity to  $w$ . There is also a partial cancellation between the two effects, and the cancellation is stronger for  $w$ , yielding a reduced net sensitivity compared to  $\Omega_m$ , by a factor of  $\approx 2$ .

For high peaks, the sign of the parameter-sensitivity is the same as for the power spectrum, but the sign reverses for low peaks, whose abundance is anti-correlated with those of high peaks. High peaks are  $\approx 2 - 3$  times more sensitive than low peaks, but there are fewer of them to help discern between models (see § 2.5). Shape noise modifies the peak function, by introducing new peaks,

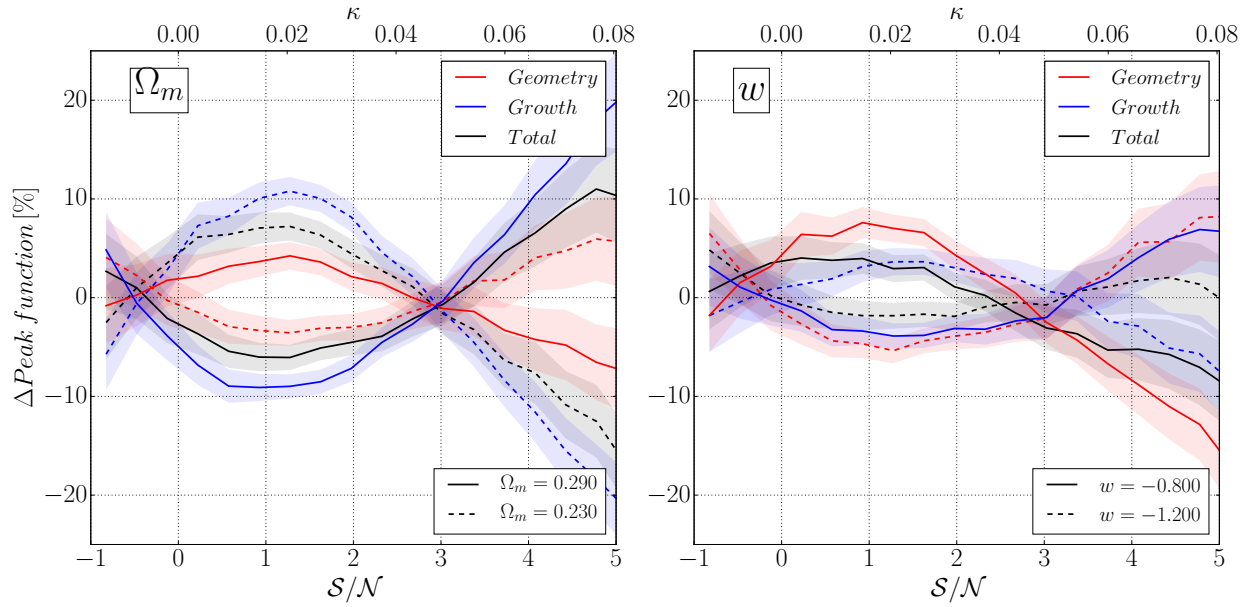


Figure 2.4: (color online) Sensitivity of peak counts to  $\Omega_m$  and  $w$  on noisy convergence maps. Both panels show the percentage difference between the peak counts in a given cosmology and in the fiducial model. Peak height is expressed in units of  $\kappa$  and in units of  $\sigma_{\text{noise}}$ ,  $S/N$ . For clarity, only two models are depicted per panel, with source galaxies at  $z_s = 1$ . The color scheme is the same as in Figs. 2.2 and 2.3. Shaded areas represent  $\pm 1$  standard deviation around the measured averages, scaled to a  $1000 \text{ deg}^2$  survey.

eliminating some, and spreading the height of those that survive from the noiseless maps. As a result, it reduces the sensitivity by a factor of  $\approx 2$ , especially for the noise-dominated low peaks, and moves the turn-over point, where the parameter-sensitivity changes sign, from  $\mathcal{S}/\mathcal{N} \approx 1$  for noiseless  $\kappa$  to  $\mathcal{S}/\mathcal{N} \approx 2.5$  ( $\mathcal{S}/\mathcal{N}$  is the height of the peaks expressed in units of  $\sigma_{noise}$ ).

For noisy  $\kappa$  and lensed galaxies at  $z_s = 2$ , the turn-over point moves to even higher  $\kappa$ , from  $\mathcal{S}/\mathcal{N} \approx 2.5$  to  $\approx 3$ , and the relative sensitivity of low peaks increases by a factor of  $\approx 2$ , while the sensitivity of high peaks remains the same.

#### 2.3.4 Minkowski functionals

Minkowski functionals (MFs) on 2D fields are topological measures on iso-contours [150]. They capture statistical information of all orders and have been shown to constrain cosmology, improving errors computed exclusively from the power spectrum, in theoretical studies [41] and also when applied to observations [49, 50].

The three MFs on a 2D map measure the area ( $V_0$ ), boundary length ( $V_1$ ) and the Euler characteristic ( $V_2$ ) of the set of points where the value of the function exceeds a pre-specified threshold ( $\kappa_{th}$ ):

$$V_0(\kappa_{th}) = \frac{1}{A} \int_{\Sigma(\kappa_{th})} da, \quad (2.8)$$

$$V_1(\kappa_{th}) = \frac{1}{4A} \int_{\partial\Sigma(\kappa_{th})} dl, \quad (2.9)$$

$$V_2(\kappa_{th}) = \frac{1}{2\pi A} \int_{\partial\Sigma(\kappa_{th})} \kappa dl, \quad (2.10)$$

where  $A$  is the total area of the map,  $\Sigma(\kappa_{th})$  is the set of points on the convergence map for which  $\kappa \geq \kappa_{th}$ , and  $\partial\Sigma(\kappa_{th})$  denotes a line integral along the curve where  $\kappa = \kappa_{th}$ . We refer the reader to [41] for a detailed description of our measurement procedure, and reproduce in Fig. 2.5 the percentage difference between the MFs for a given cosmology and the fiducial model, as a function of the threshold.

The sensitivity of all three functionals at high threshold levels is similar to that of peak counts.



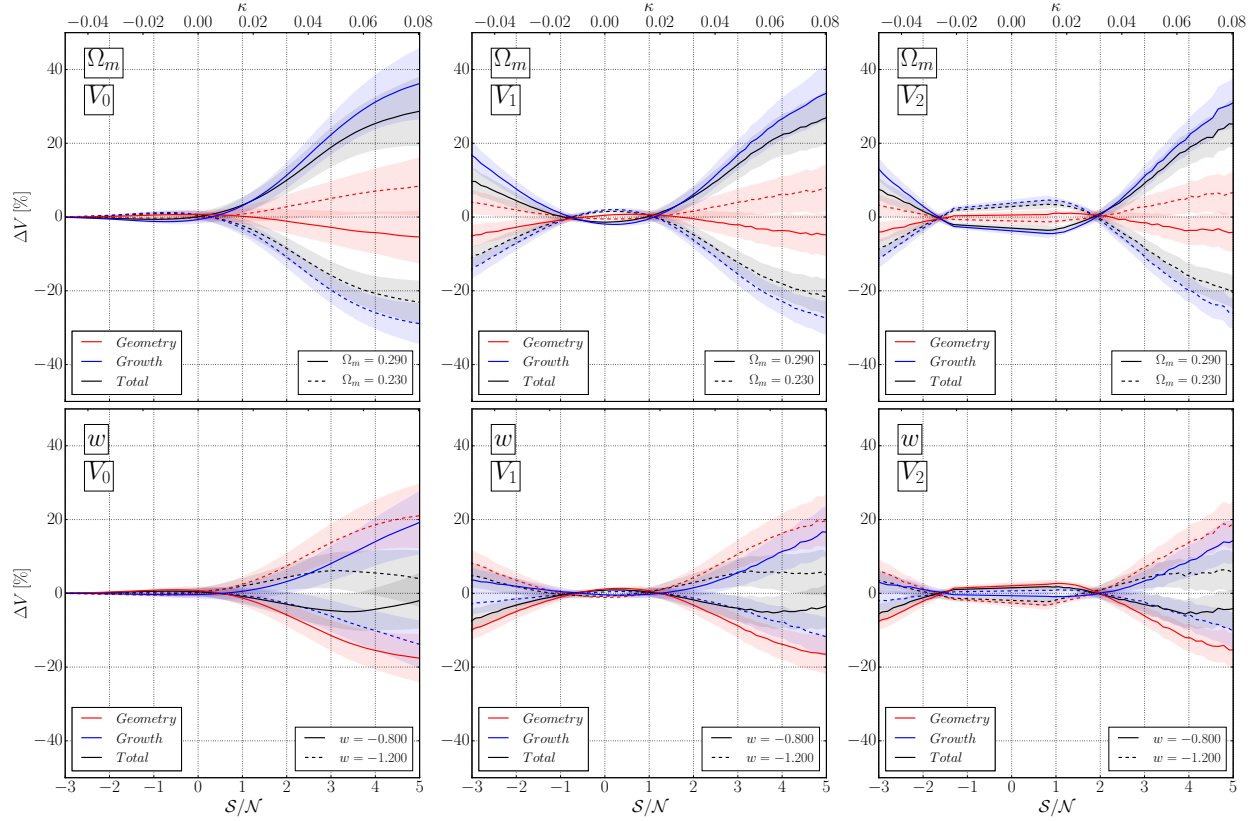


Figure 2.5: (color online) Percentage difference of the three MFs measured on noisy  $\kappa$  maps, compared to the value in the fiducial model, when changing  $\Omega_m$  and  $w$ . Left/center/right panels show the results for  $V_0/V_1/V_2$ , for noisy  $\kappa$  and source galaxies at  $z_s = 1$ . The color scheme, labeled in the legends, is the same as in Figs. 2.2-2.4. Shaded areas represent  $\pm 1$  standard deviation around the measured averages, scaled to a  $1000 \text{ deg}^2$  survey.

This is expected, since at high  $\kappa_{\text{th}}$  values, the set of points  $\kappa \geq \kappa_{\text{th}}$  increasingly coincides with the set of lensing peaks. At lower thresholds, the sensitivity of the MF is weaker, but different for each functional, suggesting that combining them should yield tighter parameter constraints.

## 2.4 Impact on parameter inference

Parameter constraints are not just determined by the sensitivity of observables, but also by their (co)variances. To assess the impact of geometry and growth on inference, we estimated the confidence levels on the parameters  $(\Omega_m, w)$  in two ways. First, we quantified how different each model is from the fiducial, using the  $\Delta\chi^2$ ,

$$\Delta\chi^2 = \sum_{i,j} \left( \mu_i - \mu_i^{\text{fid}} \right) C_{ij}^{-1} \left( \mu_j - \mu_j^{\text{fid}} \right), \quad (2.11)$$

where  $\mu_i$  is the average of an observable over the set of convergence maps for a cosmology (for instance, the binned power spectrum),  $\mu_i^{\text{fid}}$  the average for the fiducial cosmology and  $C_{ij}^{-1}$  the precision matrix. For each observable we used 20 bins, either spaced logarithmically in  $\ell$  or linearly in  $\kappa$ . We did not try to optimize the number of bins or their thresholds, since our purpose was to understand the effect of geometry and growth on the parameter uncertainties, not obtain accurate or optimal estimates for a specific survey.

We computed the precision matrix in the fiducial model, to be consistent with our calculated Fisher matrices (see below), and we corrected for its bias following [151]. The bias correction is very small,  $\approx 2\%$ , because the number of realizations used to estimate the covariance matrix ( $N_r = 1024$ ) is large compared to the dimensionality of the data vector ( $N_b = 20$ ). We scaled the results by the same factor as the error bars in the figures, so that their magnitude corresponds to what would be expected for a  $1000 \text{ deg}^2$  survey, even though in the non-Gaussian regime errors may scale logarithmically rather than as the square root of the field of view [152].

		Dependence on $\Omega_m$						Dependence on $w$									
		0.200		0.230		0.290		0.320		-0.500		-0.800		-1.200		-1.500	
		$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$
Power spectrum																	
Total	541	1550	148	421	174	444	718	1770	71	288	18	45	22	18	109	109	
Geometry-only	92	670	18	142	14	101	48	375	525	2442	92	379	102	271	532	1569	
Growth-only	839	3033	242	861	305	1110	1371	5083	528	3050	45	252	23	132	96	557	
Equilateral bispectrum																	
Total	14	38	3	13	8	8	25	41	4	5	2	3	3	4	3	5	
Geometry-only	5	9	2	5	2	4	2	10	14	47	4	12	7	8	19	40	
Growth-only	18	56	6	16	12	20	40	113	39	181	4	13	3	8	9	28	
Peak counts																	
Total	772	1120	190	266	199	232	768	825	211	399	39	48	38	26	164	93	
Geometry-only	99	336	26	76	23	70	65	223	776	1934	127	253	110	178	603	837	
Growth-only	1213	2431	317	588	361	542	1445	2071	321	931	40	114	20	83	117	373	
Minkowski functional $V_0$																	
Total	915	1153	231	272	212	265	859	976	413	828	64	90	52	56	268	194	
Geometry-only	111	455	28	107	30	81	86	281	931	2305	150	282	126	229	711	1071	
Growth-only	1464	2684	386	650	404	651	1663	2634	385	1189	38	121	26	75	116	337	
Minkowski functional $V_1$																	
Total	984	1506	245	339	229	353	901	1229	321	516	52	53	41	33	205	112	
Geometry-only	118	422	27	117	29	73	88	271	996	2595	161	285	130	214	696	1075	
Growth-only	1564	3313	400	799	423	753	1691	3043	635	2068	61	199	34	119	158	543	
Minkowski functional $V_2$																	
Total	1016	1862	255	438	253	446	997	1647	313	486	56	51	39	34	203	109	
Geometry-only	128	602	30	141	31	101	95	375	1030	3206	173	392	145	292	764	1412	
Growth-only	1613	4068	420	1000	460	997	1910	4031	736	2832	68	280	39	157	164	712	

Table 2.2:  $\Delta\chi^2$  for different cosmological models computed for the power spectrum and three non-Gaussian observables (equilateral bispectrum, peak counts and Minkowski functionals) over noisy  $\kappa$  maps with source galaxies at either  $z=1$  or  $z=2$ .

The  $\Delta\chi^2$  values are listed in Table 2.2, and are overall consistent with the conclusions from the sensitivity plots in § 2.3. The significance at which models with different  $w$ 's can be distinguished is lower than for  $\Omega_m$ , due to projection effects and the worse cancellation between geometry and growth. Geometry has stronger constraining power in  $w$  and growth does in  $\Omega_m$ , and in general the net significance is closer to that of growth than that of geometry. The observable with the lowest  $\Delta\chi^2$  is the equilateral bispectrum, especially for  $w$ , for which the cancellation between geometry and growth is particularly severe.

Even though it can strictly be used only for Gaussian-distributed data, we computed the Fisher matrix [153] for all the observables in this study, with the expectation that it provides a second-order approximation to the true parameter likelihood near its maximum:

$$F_{\alpha\beta} = \frac{1}{2}\text{Tr} [C^{-1}C_{,\alpha}C^{-1}C_{,\beta} + C^{-1}M_{\alpha\beta}], \quad M_{\alpha\beta} = \mu_{,\alpha}\mu_{,\beta}^T + \mu_{,\beta}\mu_{,\alpha}^T. \quad (2.12)$$

Here  $F_{\alpha\beta}$  is one element of the Fisher matrix, Tr stands for the trace of the matrix within brackets, the covariance is evaluated at the fiducial model and a comma denotes them partial derivative  $X_{,\alpha} \equiv \frac{\partial}{\partial\alpha}X$ . The marginalized error on a parameter is given by  $\sqrt{(F^{-1})_{\alpha\alpha}}$ , and is reported in Table 2.3. We have found the finite-difference derivatives of the covariance to be sensitive to the numeric scheme used to estimate them, especially for the bispectrum. In the case of the power spectrum and peak counts, it has been shown that this does not significantly change the parameter constraints [154, 121]. For these reasons, we have not included the cosmology-dependence of the covariance in our Fisher matrix calculations. The derivatives of the average observables were estimated using 5-point finite differences with Lagrangian polynomials.

We show the 68% confidence level contours in Fig. 2.6. The figures show that marginalized errors on  $w$  are larger than those for  $\Omega_m$  by a factor of  $\approx 15$ , and that geometry has less constraining power than growth. The confidence regions decrease when the sources are farther away, although the marginalized errors do not always do. This is due to changes in the degeneracies (i.e. the axes and tilt angles of the error ellipses). For example, the 68% contour from Minkowski functionals

	$\Delta\Omega_m$ $\times 10^3$		$\Delta w$ $\times 10^3$		$\theta$ [deg]		$FOM$	
	$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$	$z = 1$	$z = 2$
Power spectrum								
Total	14.2	4.9	269.2	127.1	-2.9	-2.0	1034	3609
Geometry-only	40.7	34.4	106.6	138.3	20.3	13.9	802	1714
Growth-only	15.2	11.6	298.8	181.5	2.9	3.6	1211	3776
Equilateral bispectrum								
Total	22.3	17.0	347.2	258.4	-0.6	1.2	131	241
Geometry-only	49.4	38.2	161.1	152.9	5.9	10.2	132	239
Growth-only	34.7	34.6	396.7	326.5	4.3	5.7	142	272
Peak counts								
Total	8.9	7.3	135.9	135.9	-3.5	-2.8	2247	2538
Geometry-only	32.9	32.9	98.3	128.5	17.9	14.2	1087	1447
Growth-only	9.4	9.8	219.2	158.3	2.4	3.5	1844	3287
Minkowski functional $V_0$								
Total	7.6	5.1	99.8	66.4	-4.0	-3.6	3259	5387
Geometry-only	29.4	36.5	89.3	146.6	17.6	13.9	1311	1425
Growth-only	4.8	4.5	115.2	79.3	2.1	3.0	3780	7018
Minkowski functional $V_1$								
Total	6.1	3.2	91.0	63.9	-3.3	-1.8	3697	6355
Geometry-only	38.0	36.6	111.6	152.5	18.4	13.4	1042	1384
Growth-only	5.1	5.6	104.1	84.1	2.5	3.6	4277	7229
Minkowski functional $V_2$								
Total	6.5	3.1	101.2	69.9	-3.3	-1.8	3436	6579
Geometry-only	36.7	40.4	106.5	159.0	18.7	14.2	1130	1489
Growth-only	5.5	6.8	109.8	99.2	2.6	3.8	4181	6962

Table 2.3: *Marginalized errors on  $\Omega_m$  and  $w$ , orientation of the Fisher ellipse (measured as the angle between its major axis and the  $w$  axis), and figure-of-merit ( $FOM$ ; defined as  $\pi/A$ , with  $A$  the area of the error ellipse). The errors correspond to a 68% confidence level, scaled to a 1000  $\text{deg}^2$  survey. All calculations were done on noisy  $\kappa$  maps with source galaxies at either  $z = 1$  or  $z = 2$ .*

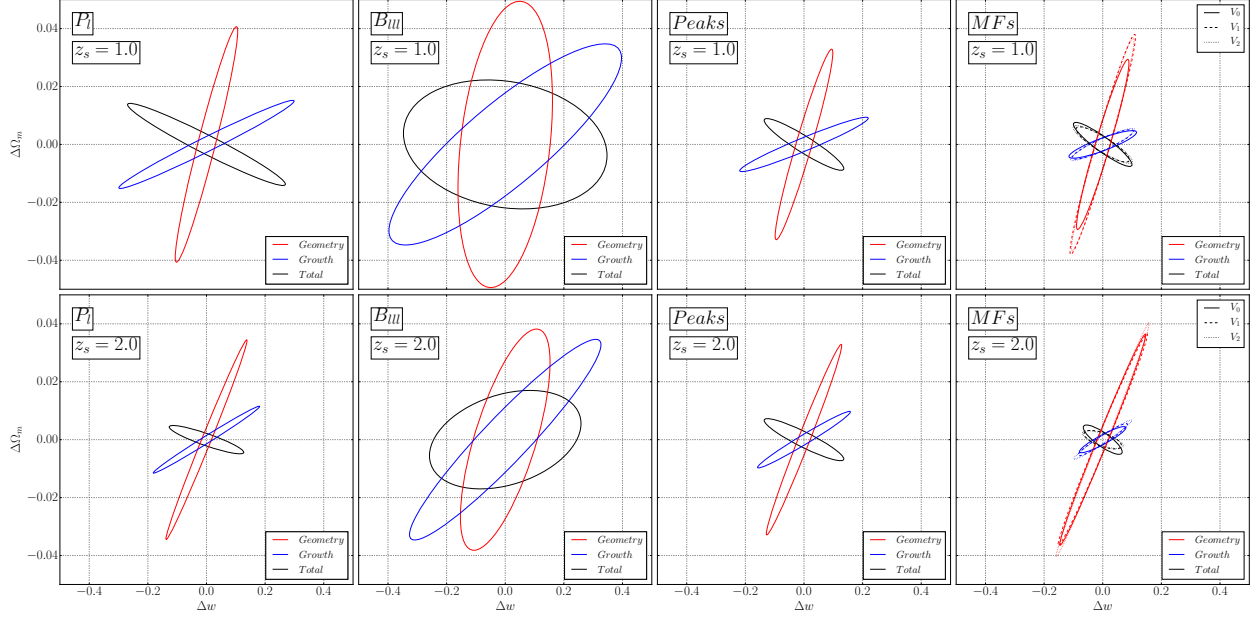


Figure 2.6: (color online) 68% Fisher error ellipses in the  $(\Omega_m, w)$  plane inferred from the power spectrum ( $P_l$ ), equilateral bispectrum ( $B_{III}$ ), lensing peaks and Minkowski functionals (MFs). Upper/lower panels show the contours for source galaxies at  $z_s = 1/z_s = 2$ . Each observable was characterized by a data vector of length 20, and the ellipses were computed neglecting the cosmology-dependence of the covariance matrix. All contours are scaled to a  $1000 \text{ deg}^2$  survey.

for geometry-only becomes more elongated and its tilt is increased towards the  $w$  axis, yielding a larger marginalized error on  $w$  for  $z_s = 2$  than for  $z_s = 1$ .

For all observables, errors on  $\Omega_m$  and  $w$  are positively correlated, when either geometry or growth is considered in isolation. For example, the geometry effect of a higher matter density is smaller comoving distances, which can also be achieved with a less negative value for  $w$ . The effect on growth of a lower DE density in the past would be a smaller suppression of gravitational collapse and a stronger gravitational field for the collapsing perturbations. The correspondingly stronger lensing signal is similar to what would be achieved with higher matter density. For the net effect, the change of the dominant effect for  $\Omega_m$  and  $w$  reverses the degeneracy direction, yielding anti-correlated errors on the parameters.

## 2.5 Discussion

The agreement between the sensitivity to  $\Omega_m$  and  $w$  of the power spectra measured on the mock  $\kappa$  maps and the analytic prediction, as well as the relative contribution of geometry and growth, validates our approach based on modified simulations.

The cancellation between geometry and growth, which further suppresses the sensitivity of WL to cosmological parameters, highlights why it is important to combine different redshift bins (tomography) to constrain DE with better precision (e.g. [155]). The suppression of the power spectrum sensitivity at small scales by galaxy shape noise highlights the importance of including other observables when analyzing weak lensing data, even if non-Gaussianities were small.

The sensitivity of the equilateral bispectrum follows a similar pattern to that of the power spectrum, but their measurement is considerably noisier, which translates into a less significant  $\Delta\chi^2$  for a given model. The addition of shape noise does not affect the mean sensitivity on small scales more than large scales, which is reasonable given the Gaussian noise model used (it does contribute to the statistical error).

We measured also the folded bispectrum, and the results are in line with those from the equilateral shape. We expect the same for all other configurations of the bispectrum, for the percentage change of the power spectrum and bispectrum does not depend on the multipole, and the cancellation between geometry and growth is a feature present at map level (see below).

The sensitivity of lensing peaks also has qualitative similarities to that of the power spectrum, but it is highly dependent on the height of the peaks. In order to assess how much of their sensitivity is a direct result of differences in the power spectrum, we computed it from Gaussian random fields (GRFs) built with the same power spectra as the  $\kappa$  maps generated through ray-tracing. The result of this exercise is shown in Fig. 2.7. We have found that the sensitivity of low peaks is reduced by a factor of  $\approx 2$ , and the sensitivity of the high peaks increases (although there are fewer high peaks in the GRFs). Overall, the  $\Omega_m$ -sensitivity of the counts cannot be fully explained by the power spectrum.

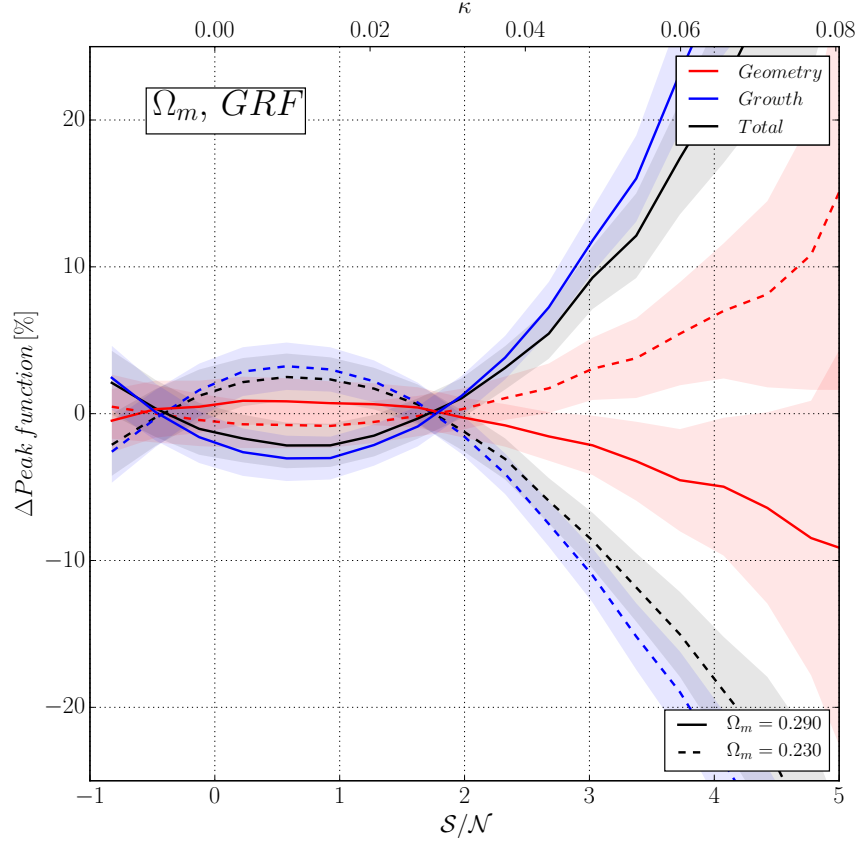


Figure 2.7: (color online) Sensitivity of lensing peak counts to  $\Omega_m$ , derived from a set of Gaussian random fields with the same power spectra as that measured on noisy convergence maps from large-scale structure. Shaded areas represent 1 standard deviation errors in a  $1000 \text{ deg}^2$  survey. Compare with the left panel of Fig. 2.4.



To better understand the origin of the dependence of the peak counts' sensitivity to peak height, we look at the 3D dark matter halo counts. It is natural to compare these quantities, since high peaks have long been known to be strongly correlated with individual high-mass DM halos hosting galaxy clusters [156, 157, 158]. The average number of halos of a given mass to a fixed redshift per solid angle can be expressed as an integral of the product of the volume element (geometry) and the halo mass function (growth).

$$\frac{dn}{d\ln M d\Omega}(M) = \int_0^{z_s} dz \frac{dV}{dz d\Omega}(z) \frac{dn}{d\ln M}(z, M) \quad (2.13)$$

We have computed the contribution from each effect as a function of halo mass, and displayed the results in Fig. 2.8. The sensitivity for halo masses above  $\approx 10^{12} h^{-1} M_\odot$  tracks that of high peaks, but this is not the case for low peaks / lower mass halos. This is in agreement with previous studies that showed a link between high peaks and single high-mass halos, while finding that lower peaks are associated instead with constellations of 4-8 low-mass halos at a range of redshifts [159]; a similar peak-halo correlation has been seen in recent CFHTLenS data [160]. High peaks then seem to measure, like halos, measure a combination of growth and the volume element.

The sensitivity of the low peak counts does not track that of halo counts; but these peaks are important for cosmology. When normalized by the standard deviation for the fiducial model, the difference in peak counts from the fiducial model has a maximum in the low significance region (see Fig. 2.9). Low peaks have also been found to contribute to cosmological parameter constraints more than high peaks, which is in agreement with previous studies [159, 121], including an analysis of peak counts in the the CFHTLenS data [47].

The sensitivity of the Minkowski functionals, as well as its decomposition into geometry and growth effects, qualitatively traces that of lensing peaks, especially at high  $\kappa$  levels.

Finally, the fact that we observe a partial cancellation between geometry and growth, especially when changing  $w$ , in all the statistics and topological descriptors analyzed, suggests that this property is present already at the map level. In order to investigate whether this is the case,

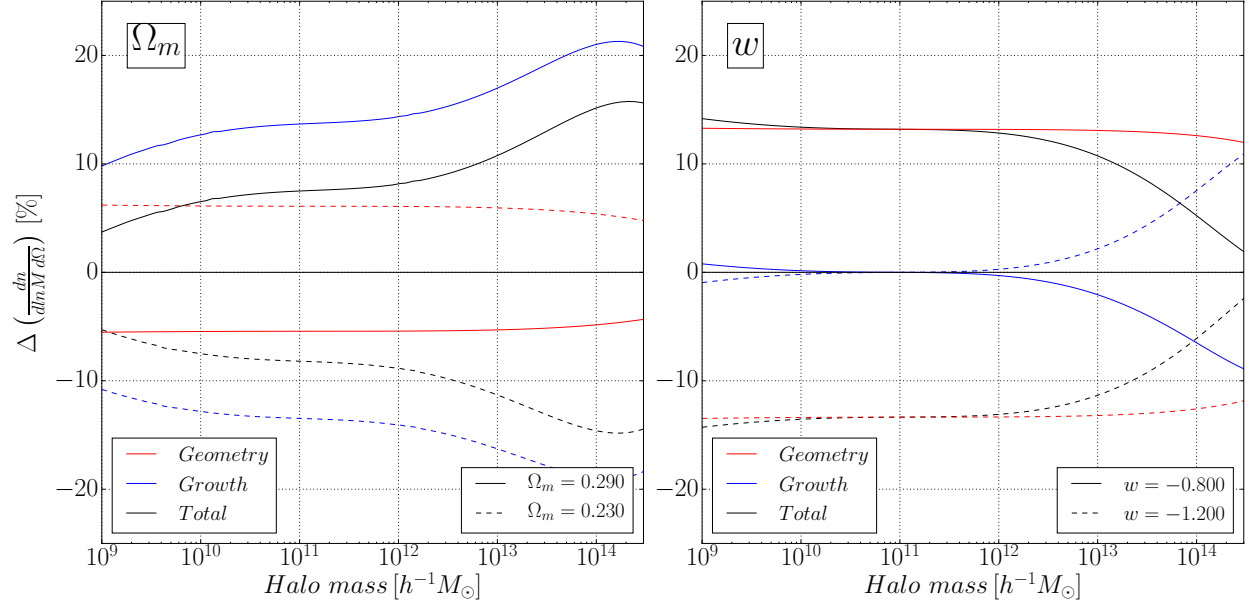


Figure 2.8: (color online) Sensitivity of DM halo abundance to  $\Omega_m$  (left panel) and  $w$  (right panel). The percentage difference in the total number of halos per unit solid angle to  $z = 1$  between a model and the fiducial cosmology, as a function of the halo mass. The net effect (black) is decomposed into its geometry (red) and growth (blue) components.

we have examined the difference-maps between each model and the fiducial, including either the geometry or growth effect alone. These maps are shown in Fig. 2.10 for the model with  $w = -1.2$ . The modified angular positions of structures in the maps built including each effect, due to different ray deflections, prevent us from <https://www.overleaf.com/project/5eb376fe552dce0001998108> directly demonstrating a cancellation of the lensing signal by adding these maps together. Nevertheless, the geometric and growth-induced distortions in the two panels of Fig. 2.10 clearly show the same structures at roughly the same locations, but with the sign of their  $\Delta\kappa$  values reversed. We conclude that the geometry vs growth cancellation indeed is a property at the map level and we therefore expect it to affect any observable, including those not analyzed here.

## 2.6 Conclusions

We have validated the use of N-body simulations and ray-tracing to separately study the effect of geometry and growth on weak lensing observables. This allows us to extend past analyses to

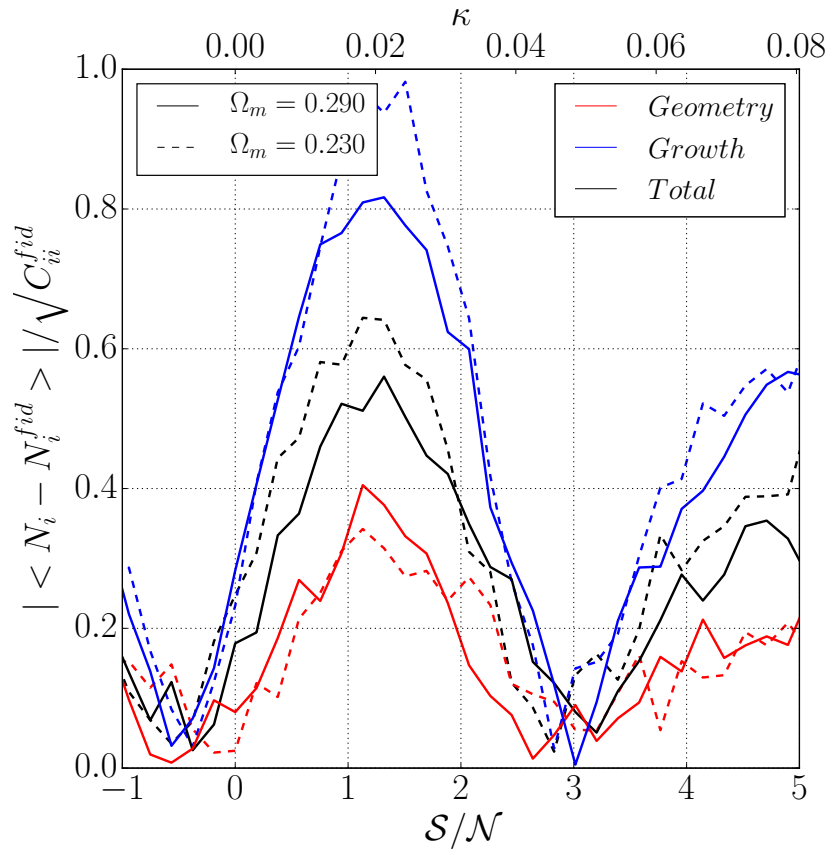


Figure 2.9: (color online) *Difference in number of peaks from the fiducial cosmology, normalized by the standard deviation in the fiducial model, for a 1000 deg<sup>2</sup> survey.*

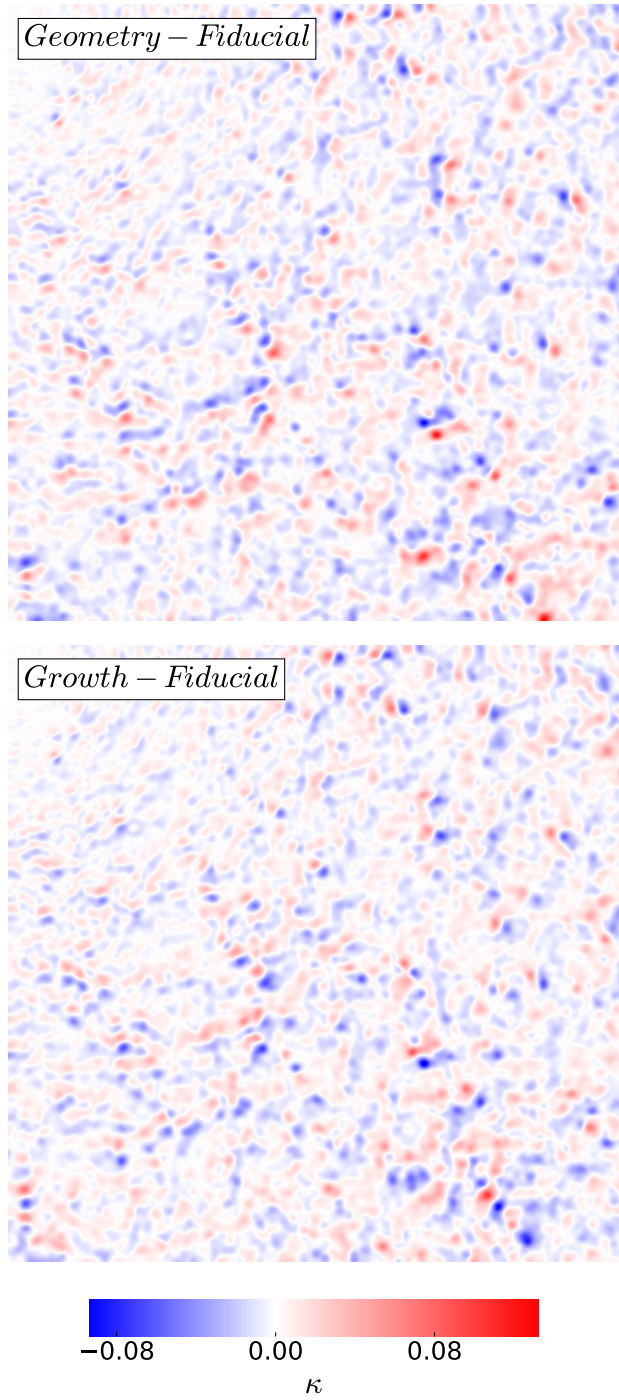


Figure 2.10: (color online) Difference  $\Delta\kappa$  between a single realization of the noisy  $\kappa$  field generated including only geometry (top panel) or growth effects (bottom panel) for  $w = -1.2$ , and the corresponding realization in the fiducial model ( $w = -1.0$ ).

non-Gaussian statistics and topological features that do not admit a simple analytic treatment.

Our analysis confirms that the sensitivity of non-Gaussian observables to cosmology shares some characteristics with that of the power spectrum. They suffer a partial cancellation between geometry and growth on top of the loss of sensitivity due to integrating (projection) effects. This cancellation is more severe for  $w$ , reducing even further the sensitivity of WL to that parameter compared to  $\Omega_m$ .

Galaxy shape noise dominates the power spectrum at high multipoles, reinforcing the case to use alternative observables to analyze weak lensing data on small scales. The bispectrum has higher statistical noise, but shape noise does not suppress its average sensitivity at high multipoles as it does for the power spectrum. The lensing peaks' sensitivity is highly dependent on the peak height, with high peaks tracking the behavior of dark matter halo counts, but low peaks having an important influence on parameter constraints. The sensitivity of Minkowski functionals is similar to that of peak counts, which is not surprising at high  $\kappa$  levels. The similarities between statistics, such as the cancellation of geometry and growth effects, arises from the fact that this property is present at map level.

The partial cancellation, together with projection effects, yields weak constraints for  $w$ , and underscores the need to combine information from different redshifts to tighten constraints on DE. Marginalized errors on  $\Omega_m$  and  $w$  are similar to those calculated from growth-only effects. This suggests that combining WL data with probes that strongly constrain the expansion history through geometry, such as BAO measurements, may be especially beneficial to tighten constraints.

## Chapter 3: Do dark matter halos explain lensing peaks?

### 3.1 Introduction

Weak gravitational lensing (WL) of background sources by large-scale structure (LSS) is a promising technique to study dark matter (DM) and dark energy (DE) [3] as a consequence of its sensitivity to both structure growth and the expansion history of the universe. Ongoing and future surveys such as the Dark Energy Survey (DES<sup>1</sup>), the Euclid Mission<sup>2</sup>, the Wide Field Infrared Survey Telescope (WFIRST<sup>3</sup>) and the Large Synoptic Survey Telescope (LSST<sup>4</sup>) will deliver WL datasets with unprecedented precision, sky coverage and depth. For a comprehensive treatment of weak lensing in a cosmological context, we refer the reader to the following reviews [4, 161, 5].

On small scales, WL probes the matter density field in the non-linear regime, independent of the matter's nature or dynamic state. Thus, in order to optimally extract cosmological information from the upcoming WL surveys, we need observables that go beyond quadratic statistics such as the two-point correlation function or its Fourier transform, the power spectrum. Various strategies have been proposed to capture non-Gaussian information, from the use of higher-order moments and correlation functions such as the bispectrum [129, 130, 162, 163], to the adoption of topological features from WL maps such as Minkowski functionals [41, 50] or peak counts [132].

Lensing peaks, defined as local maxima of the convergence or shear field, are particularly simple to extract from mass-aperture maps, and have been shown to constrain cosmology both theoretically [31, 32, 164] and, recently, in observations [47, 48, 165]. Peaks are usually classified based on their absolute height or significance level, defined as their signal-to-noise ratio ( $S/N$ ), the noise being caused by our imperfect knowledge of the intrinsic shapes of the background

---

<sup>1</sup><http://www.darkenergysurvey.org>

<sup>2</sup><http://sci.esa.int/euclid/>

<sup>3</sup><http://wfirst.gsfc.nasa.gov>

<sup>4</sup><http://www.lsst.org>

galaxies.

Peak counts are also special because their physical origin and sensitivity to cosmology can, in principle, be understood and related to specific structures of the cosmic web. While our understanding is not yet complete, it is clear that halos are important contributors to peak counts. Shear peaks were initially considered for cluster selection, and the connection of high-significance peaks ( $S/N > 4 - 5$ ) to single massive halos has been established in the literature [156, 157, 158]. Lower-significance peaks are typically associated with constellations of lower-mass halos [159, 160] and contribute significantly to the cosmological information in convergence maps [32, 159].

Predicting analytically the abundance of peaks is difficult, as it depends on projections of non-linear structures. N-body simulations can predict peak counts at a high computational cost that will only increase with the high volumes required by upcoming WL surveys. The need to predict not only the peak number density but also its covariance would further raise the total cost. The halo-peak connection has inspired some models that would circumvent the need for full N-body simulations by using either analytical models based on Gaussian random fields [166, 167, 168] or stochastic fast simulations based on the halo model [169, 170]. This could prove extremely useful by reducing the computational requirements for N-body simulations by 2-3 orders of magnitude.

The main goal of this work is to assess the validity of halo-based models for cosmological parameter inference. In particular, we compare results from full N-body simulations with those of a recent publicly available algorithm, CAMELUS [170]. In previous work [170], this model was found to predict accurately peak counts from N-body simulations for a specific cosmology. Here, we expand the comparison of peak counts to a wide range of different cosmologies, and also examine their predicted covariance matrices, showing how differences affect the resulting parameter credibility regions. We also review the importance of the cosmology-dependence of the covariance matrix in the context of precision parameter inference [154].

The rest of the paper is organized as follows. In Sec. 3.2 we describe the methods used to predict peak counts using N-body simulations and CAMELUS, and infer constraints for cosmological parameters. In Sec. 3.3 we show how both models compare in terms of peak counts, covariance

Parameter	Symbol	Value
Matter density	$\Omega_m$	0.260
Dark energy density	$\Omega_{\text{DE}}$	$1.0 - \Omega_m$
Amplitude of fluctuations at $8 h^{-1}\text{Mpc}$	$\sigma_8$	0.800
Hubble constant	$h$	0.72
Dark energy eq. of state	$w$	-1.0
Scalar spectral index	$n_s$	0.96
Effective number of relativistic d.o.f.	$n_{\text{eff}}$	3.04

Table 3.1: Cosmological parameters for the fiducial model. All other cosmologies share these parameters except  $\Omega_m$  and  $\sigma_8$ .

matrices, and credible contours. We then discuss our main findings (Sec. 3.4), identifying potential origins for the differences between the two models and how CAMELUS could be modified to match N-body predictions more accurately. Our main conclusions are summarized in Sec. 3.5.

### 3.2 Predicting peak counts

We generated convergence maps for a suite of 162 flat  $\Lambda$ CDM cosmologies covering the  $\{\Omega_m, \sigma_8\}$  plane using both N-body simulations and CAMELUS. Table 3.1 presents the cosmological parameters for our fiducial cosmology, which are consistent with the 9-year Wilkinson Microwave Anisotropy Probe (WMAP) results [171] for ease of comparison with past simulation efforts.

We sampled the parameter space with a modified latin hypercube algorithm implemented in the publicly available lensing package LENSTOOLS [143], and based on a coordinate transformation that converts a randomly sampled rectangle into an ellipse:

$$(r, \phi) \rightarrow (x = ar^n \cos \phi, y = br^n \sin \phi) \quad (3.1)$$

with  $(r, \phi) \in [0, 1] \times [0, 2\pi]$ . We adjusted the semi-axes  $a$  and  $b$  so that the region explored covered all areas with a significant likelihood according to past WL peak counts studies [31]. We centered the ellipse on our fiducial cosmology ( $\Omega_m = 0.260, \sigma_8 = 0.800$ ), and rotated it so that



its semi-major axis became parallel to the direction of maximum degeneracy between the two parameters. The exponent  $n$  controls the sampling concentration, with  $n > 1/2$  yielding samples whose density grows towards the center of the ellipse. We used  $n = 3/2$ .

Based on the likelihood estimated from a first batch of 100 cosmologies, we added manually 62 cosmologies in sparsely sampled regions, such as the contours' tails. Doing so reduced the sampling error in the likelihoods, as discussed in Sec. 3.4.

### 3.2.1 N-body simulations

Our simulation pipeline is described in detail in [143]. For each cosmology, we evolved a single  $(240 h^{-1}\text{Mpc})^3$  volume with GADGET2 [144], large enough to cover the intended  $3.5 \times 3.5 \text{ deg}^2$  field of view to a distance beyond the lensed sources' redshifts. Every simulated box contains  $512^3$  DM particles, which yields a mass resolution of  $M_p \approx 10^{10} M_\odot$ . All lensed source galaxies were placed at a redshift of  $z_s = 1$ , and  $80 h^{-1}\text{Mpc}$  thick lens planes were stacked between the galaxies and the observer. Each lens plane is the result of slicing a snapshot along a coordinate axis, and applying to it a random shift and rotation, allowing us to generate 500 independent realizations from a single N-body run. Lens planes were converted to potential planes and a multi-plane ray-tracing algorithm was used to generate  $1,024 \times 1,024$  pixels convergence maps with a pixel size of  $\approx 0.2 \text{ arcmin}$ . We used a higher resolution for the potential planes,  $4,096 \times 4,096$  pixels, to avoid a loss of power on small scales [159]. We deployed and managed the simulations and their output using LENSTOOLS [143].

Since the unperturbed galaxy shape is unknown, we accounted for an intrinsic ellipticity noise following [147] and added a 2-D Gaussian random noise with zero mean and standard deviation

$$\sigma_{pix} = \sqrt{\frac{\sigma_\epsilon^2}{2n_g A_{pix}}} \quad (3.2)$$

with intrinsic ellipticity  $\sigma_\epsilon = 0.4$  as in [170], a galaxy density of  $n_g = 25 \text{ arcmin}^{-1}$  and pixel area defined by the field-of-view and map resolution. We smoothed the noiseless and noise-only

maps applying a Gaussian filter with a characteristic width of  $\theta_G = 1$  arcmin –see Eq. 3.3– before combining them, and extracted their local maxima, recording them in the form of peak catalogues.

$$W(\theta) = \frac{1}{\pi\theta_G^2} \exp\left(-\frac{\theta^2}{\theta_G^2}\right) \quad (3.3)$$

### 3.2.2 Camelus

CAMELUS is a halo-based model that generates fast stochastic simulations of convergence maps. Instead of evolving the matter density field from high redshift dynamically, it assumes that halos are the primary contributors to the lensing signal and discretizes the space between the lensed galaxies and the observer in redshift bins, populating them with halos whose masses are sampled from an analytical function [172]. Each halo follows a Navarro-Frenk-White (NFW) density profile [173] and is placed randomly within its redshift bin. We refer the reader to [170] for an in-depth description of the model.

We ran CAMELUS for each of the same set of 162 cosmologies as with the N-body simulations, generating 500 independent realizations in each case. The resulting smoothed, noiseless convergence maps were combined with shape noise that is statistically the same as the one used with the N-body maps, and their peaks extracted with the same routines. The values we used for the relevant tunable parameters in CAMELUS are given in Table 3.2.

### 3.2.3 Parameter inference

Bayes’ theorem relates the probability distribution for a set of cosmological parameters, given an observation, to the likelihood of the observed data given values for those parameters

$$p(\theta|\mathbf{x}^{obs}, M) = \frac{p(\mathbf{x}^{obs}|\theta, M)p(\theta, M)}{p(\mathbf{x}^{obs}, M)} \quad (3.4)$$

Parameter	Symbol	Value
Field of view	$f_{ov}$	$210.0 \times 210.0 \text{ arcmin}^2$
Pixel size	-	0.205 arcmin
Smoothing scale	$\theta_G$	1.0 arcmin
Minimum halo mass	$M_{min}$	$10^{11} h^{-1} \text{M}_\odot$
Maximum halo mass	$M_{max}$	$10^{17} h^{-1} \text{M}_\odot$
Maximum halo redshift	$z_{max}$	1.0
No. of redshift bins	$n_z$	10
Halo profile inner slope	$\alpha$	1.0
Halo concentration (norm.)	$c_0$	11.0
Halo concentration (slope)	$\beta$	0.13
Galaxies redshift	$z_{gal}$	1.0
Galaxy density	$n_g$	$25.0 \text{ arcmin}^{-2}$
Ellipticity noise	$\sigma_\epsilon$	0.4

Table 3.2: The main tunable parameters of CAMELUS and their values used in this study. Dark matter halos are assumed to have a Navarro-Frenk-White (NFW) density profile, defined by its inner slope ( $\alpha$ ), and its concentration parameter, the ratio between the virial and scale radii, determined itself by  $c_0$  and  $\beta$ :  $c_{\text{NFW}} \equiv \frac{c_0}{1+z} \left( \frac{M}{M_\star} \right)^\beta$ , where  $z$  is the halo's redshift,  $M$  its mass and  $M_\star$  its pivot mass (see [170] for a detailed description of the halo density profile characterization).

where  $p$  is the probability,  $\theta$  represents the set of parameters that determine the model  $M$  and  $\mathbf{x}^{obs}$  is a data vector that depends on observations. Throughout this study we assume  $\Lambda$ CDM is a correct description of the universe, hence the evidence (denominator) acts just as a normalizing factor and we can drop the implicit dependence on the model. We use a non-zero prior within the parameter region that we explore, and zero outside:

$$p(\theta|\mathbf{x}^{obs}) \propto p(\mathbf{x}^{obs}|\theta) \equiv \mathcal{L}(\theta) \quad (3.5)$$

Our observable is the peak function defined as the peak counts binned by their height or significance level ( $S/N$ , height in units of the r.m.s. ellipticity noise).

If we assume that our observable follows a multivariate Gaussian distribution, its log-likelihood, up to an additive constant, has the form:

$$L_{vg} = \ln \left[ (2d)^d \det C(\theta) \right] + \Delta \mathbf{x}^T(\theta) \widehat{C^{-1}}(\theta) \Delta \mathbf{x}(\theta) \quad (3.6)$$

where  $\Delta \mathbf{x}$  is the difference between the mean peak function in each cosmology from its value in the fiducial ( $\Omega_m = 0.260, \sigma_8 = 0.800$ ) cosmology, and  $\widehat{C^{-1}}$  is the precision matrix (the inverse of the covariance matrix), estimated from the data. We follow the same notation as [154], and call it  $L_{vg}$ ,  $L$  indicating it is a "log-likelihood",  $v$  that it includes a "varying" (i.e. cosmology-dependent) covariance matrix, and  $g$  that the assumed model is "Gaussian".

Means and covariance matrices are computed from the  $N = 500$  realizations available in each cosmology:

$$\Delta \mathbf{x}(\theta) = \bar{\mathbf{x}}(\theta) - \bar{\mathbf{x}}(\theta_{fid}) \quad (3.7)$$

$$C(\theta) = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i(\theta) - \bar{\mathbf{x}}(\theta))(\mathbf{x}_i(\theta) - \bar{\mathbf{x}}(\theta))^T \quad (3.8)$$

In many cases, evaluating the covariance matrix at each point of the parameter space becomes computationally too expensive, and a constant covariance is used instead. As in [154], we assess the effect of this simplification by evaluating two approximations to the full Gaussian likelihood. The first is to use a "semi-varying" covariance matrix; i.e., we let the covariance matrix change with cosmology within the  $\chi^2$  term but not the determinant term in Eq. 3.6. Following the notation in [154] we call it  $L_{svg}$ . The second is to compute the likelihood with a "constant" covariance matrix, evaluated at the fiducial model, in all terms. We call this  $L_{cg}$ :

$$L_{svg} = \Delta \mathbf{x}^T(\theta) \widehat{C^{-1}}(\theta) \Delta \mathbf{x}(\theta) \quad (3.9)$$

$$L_{cg} = \Delta \mathbf{x}^T(\theta) \widehat{C^{-1}}(\theta_{fid}) \Delta \mathbf{x}(\theta) \quad (3.10)$$

Note that the precision matrices in Eqs. 3.6, 3.9 and 3.10 have a "hat" on top, while the covariance matrix in Eq. 3.8 does not. That is because the inverse of a covariance matrix estimated from data is not an unbiased estimator for the precision matrix. There are two ways to correct for the bias.

The most common [151] is to re-scale the inverse of the estimated covariance matrix:

$$\widehat{C^{-1}} = \frac{N - d - 2}{N - 1} C^{-1} \quad (3.11)$$

where  $N$  is the number of realizations per cosmology (500 in our case) and  $d$  the dimension of the observable (number of bins in the peak function).

An alternative approach is to use a non-Gaussian likelihood, as described in [174]. In this case we can also use a constant or varying covariance matrix in each of the log-likelihood terms and, following the same notation, drop the  $g$  subscript since the model is not a Gaussian anymore. The functional form for these models is as follows:

$$L_v = \ln \left[ \frac{\det C(\theta)}{c_p^2} \right] + N \left[ 1 + \frac{\Delta \mathbf{x}^T(\theta) C^{-1}(\theta) \Delta \mathbf{x}(\theta)}{N-1} \right] \quad (3.12)$$

$$L_{sv} = N \left[ 1 + \frac{\Delta \mathbf{x}^T(\theta) C^{-1}(\theta) \Delta \mathbf{x}(\theta)}{N-1} \right] \quad (3.13)$$

$$L_c = N \left[ 1 + \frac{\Delta \mathbf{x}^T(\theta) C^{-1}(\theta_{fid}) \Delta \mathbf{x}(\theta)}{N-1} \right] \quad (3.14)$$

$$(3.15)$$

with a normalizing factor

$$\bar{c}_p = \frac{\Gamma\left(\frac{N}{2}\right)}{[\pi(N-1)]^{d/2} \Gamma\left(\frac{N-d}{2}\right)} \quad (3.16)$$

where  $\Gamma$  is the usual Gamma function and  $N > d$ .

In the limit  $N \gg d$  both methods are equivalent. We used peak functions with a relatively small number of bins (see below) compared with the number of realizations per model and there were no discernible differences between the credible contours generated using the two approaches.

For inference, we decided to use few bins in the peak function so that covariance bias is not an issue. We set an edge at  $\mathcal{S}/\mathcal{N} = 3.0$ , the threshold below which peak counts are dominated by noise. This allowed us to separate clearly analyses done with only high-significance peaks (as in [154]) from analyses also including low-significance and even negative peaks. The upper and

Observable	$S/N$ bins
$\mathbf{n}_{\text{pk}}^{(100)}$	100 equally-sized bins in $[-2.0, \dots, 6.0]$
$\mathbf{n}_{\text{pk}}^{(10)}$	$[-\infty, -1.0, 0.0, 1.0, 2.0, 3.0,$ $3.5, 4.0, 4.5, 5.0, +\infty]$
Model	$\bar{\mathbf{n}}_{\text{pk}}^{(10)}$
N-body	$[23.8, 292.5, 1125.7, 1457.3, 735.4,$ $130.5, 59.8, 27.8, 13.3, 17.7]$
CAMELUS	$[15.3, 255.6, 1145.8, 1535.1, 721.7,$ $113.3, 48.1, 21.2, 10.4, 15.7]$

Table 3.3: Description of the thresholds used in this study to bin the convergence peak counts by their signal-to-noise ( $S/N$ ) ratio, as well as the mean peak counts from data obtained from both the N-body and the CAMELUS models in the fiducial cosmology, in the bins used for inference.

lower  $S/N$  edges were chosen to avoid the rejection of models due to the presence of empty bins with their corresponding singular covariance matrices. We also ensured that there are at least 10 peaks from the fiducial cosmology in the bin with the lowest number and defined the 10-bin peak function described in Table 3.3,  $\mathbf{x}^{obs} \equiv \mathbf{n}_{\text{pk}}^{(10)}$ , as the observable for this study. We did not optimize the bins' edges to maximize the predictive power of the models.

Table 3.3 also displays  $\mathbf{n}_{\text{pk}}^{(100)}$ , a peak function with 100 equally spaced bins that was used to highlight differences in peak counts from the two models.

We are forced to interpolate for all the  $(\Omega_m, \sigma_8)$  combinations not found in our collection of simulated cosmologies in order to compute smooth credible contours. Our interpolation grid covers the region  $\Omega_m \in [0.160, 0.600]$  and  $\sigma_8 \in [0.150, 1.250]$  with a resolution of 0.001 on each axis. Within that region we know that our sample reproduces  $2\sigma$  (95.4%) contours from CAMELUS within 20% –see Sec. 3.4–, and we verified that a finer grid did not change the results.

Interpolating peak counts is straightforward, and can be done when using a constant covariance to calculate the likelihood, but becomes problematic when an estimation for the covariance matrix is also needed. We interpolated the log-likelihood instead, and used a linear model because its results are easy to interpret, it does not require any tunable parameter like smoothing, and it does not introduce any spurious high-likelihood values from fitting high-order polynomials. We verified

that our results do not change when using a different interpolator, such as radial basis functions; this agrees with the findings in previous studies such as [165].

### 3.3 Results

Our main results are the comparison between the two models regarding peak counts, covariance matrices and credible contours, together with the impact of using a cosmology-dependent covariance for inference.

Fig. 3.1 shows mean peak counts as a function of their height, with and without galaxy shape noise, for three representative cosmologies that are characterized by the degeneracy parameter defined as in [165],  $\Sigma_8 \equiv \sigma_8 \left( \frac{\Omega_m}{0.3} \right)^{0.6}$ . In each cosmology, we calculated the average of the peak function,  $\bar{n}_{\text{pk}}^{(100)}$ , over 500 smoothed maps generated with the two models. We did this before and after adding noise as described in 3.2.1. Noiseless maps from N-body simulations exhibit up to 50% fewer peaks around the maximum of the distribution at  $\mathcal{S}/\mathcal{N} \sim 0$ , with higher counts in the tails. Nevertheless, the two models agree well for peaks with  $\mathcal{S}/\mathcal{N} > 1$ , which constrain cosmology the most (see below). The addition of noise dilutes the differences for low-significance counts, especially for cosmologies with small  $\Sigma_8$ , and has the opposite effect for high-significance peaks, with N-body noisy maps yielding more counts for  $\mathcal{S}/\mathcal{N} > 3$ , especially for cosmologies with high  $\Sigma_8$ .

As a global measure of how different the peak histograms from the two models are, we integrated the area between them, divided the result by the surface under the N-body histograms –see Eq. 3.17–, and plotted it as a function of  $\Sigma_8$  in Fig. 3.2.

$$\Delta(\%) \equiv \frac{\sum_{i=1}^{N_{\text{bins}}} |N_{\text{peaks}}^i_{\text{CAMELUS}} - N_{\text{peaks}}^i_{N\text{-body}}|}{\sum_{i=1}^{N_{\text{bins}}} N_{\text{peaks}}^i_{N\text{-body}}} \quad (3.17)$$

Counts from noiseless convergence maps generated with the halo-based model are in better

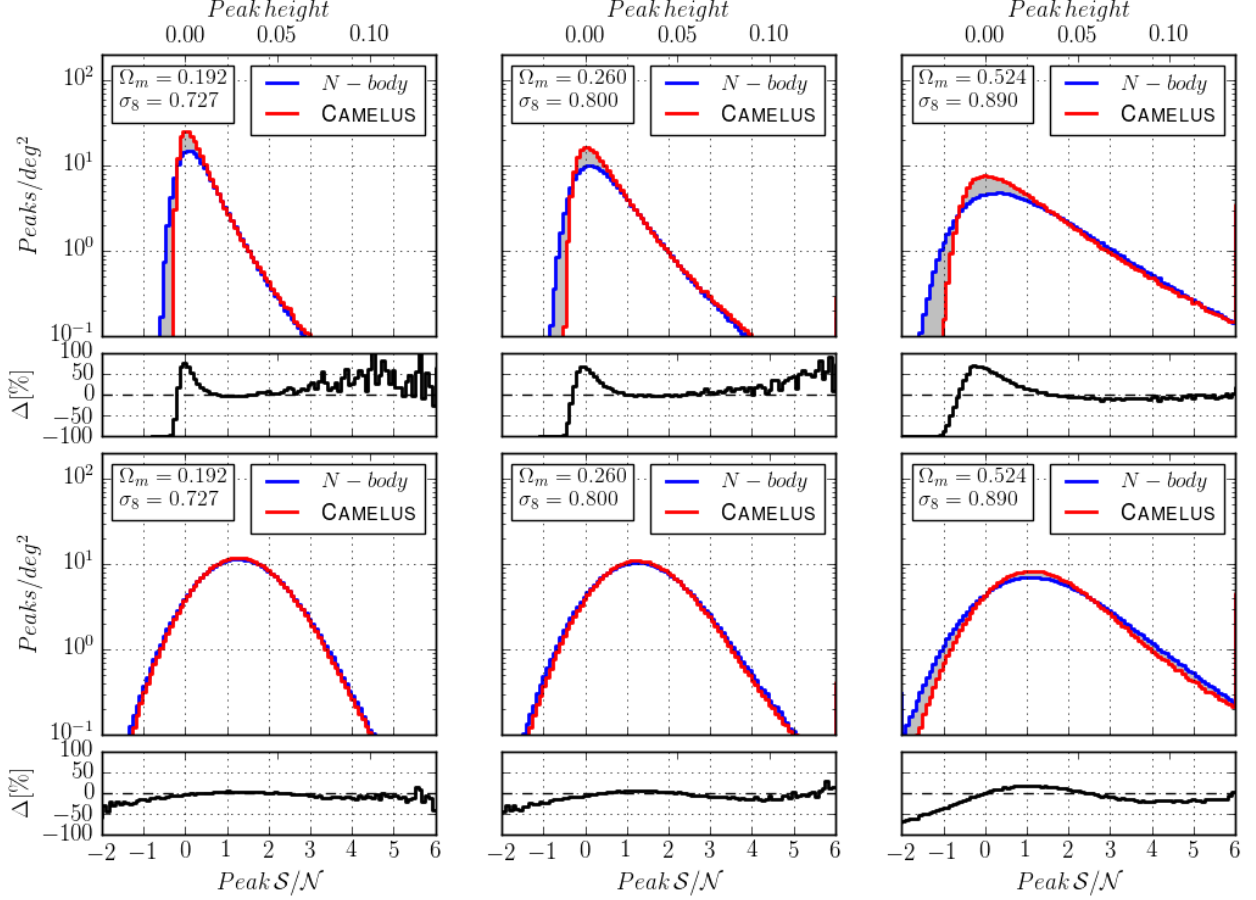


Figure 3.1: Comparison of mean peak counts as a function of their height between N-body simulations (blue) and CAMELUS (red). Counts are normalized to  $1 \text{ deg}^2$  of sky and height is expressed in absolute value and as a signal-to-noise ratio ( $S/N$ ). The upper panels show the results from smoothed convergence maps without shape noise; the lower panels add shape noise. Three different cosmologies are displayed with increasing parameter  $\Sigma_8$  from left to right (0.556, 0.734 and 1.244). In black, we show the fractional difference between the two models ( $\Delta[\%] \equiv (N_{\text{CAMELUS}} - N_{N\text{-body}})/N_{N\text{-body}}$ ), and the area between the histograms is shaded. Adding noise reduces the discrepancies between the models but the effect depends on cosmology. While the discrepancies are almost erased for cosmologies with small  $\Sigma_8$ , for the rest N-body data yield lower counts near the maximum of the distribution and higher counts in the tails. The differences grow with  $\Sigma_8$ .



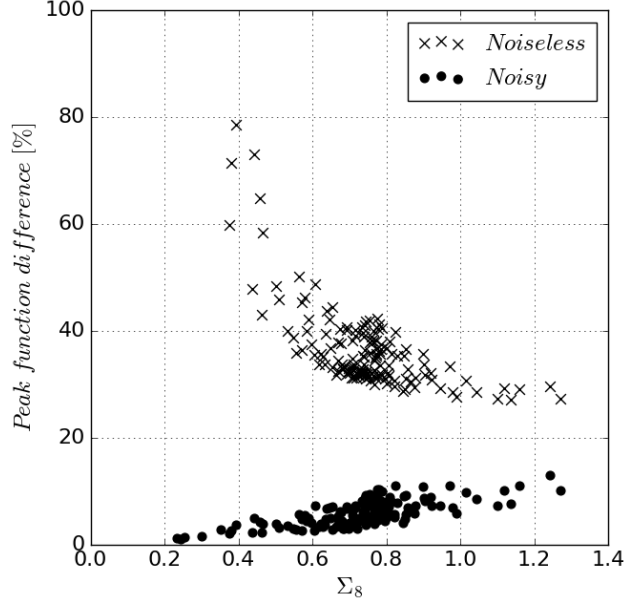


Figure 3.2: Global comparison of peak counts. For each cosmology, the area between the N-body and CAMELUS histograms as a percentage of the area enclosed by the N-body histogram (Eq. 3.17) is plotted against  $\Sigma_8$ . Differences from noiseless maps (crosses) are significantly reduced by adding noise (dots), so that the difference stays below 20% in all cases. The reduction is more important for cosmologies with small  $\Sigma_8$ , for which noise dominates.

agreement with those from N-body simulations as  $\Sigma_8$  increases, pointing to a higher non-halo contribution to peaks for small  $\Sigma_8$  cosmologies. Adding noise reduces the global differences to less than 20% in all cases. As expected, the reduction is stronger for cosmologies with small  $\Sigma_8$  where peak counts are dominated by noise. Thus, the agreement between models worsens as  $\Sigma_8$  increases.

Calculating the likelihood of a cosmological model needs an estimate of the covariance matrix, as seen in Sec. 3.2.3. We analyzed the covariances for  $\mathbf{n}_{\text{pk}}^{(10)}$ , the data vector used to draw the credible contours. Fig. 3.3 shows this comparison for the fiducial cosmology. Specifically, we display the correlation matrices after substituting their diagonal terms with the variances divided by the mean peak counts. These normalized matrices allow for a comparison of the variance and correlations for each bin, irrespective of its mean peak count.

N-body data yield higher absolute values in all matrix elements. Positive and negative peaks have higher correlations among themselves, while being anti-correlated against one another. CAMELUS

data, on the other hand, gives weakly anti-correlated peak counts with a smaller variance. The weak anti-correlation in the CAMELUS data can be attributed to the condition that the total mass in all halos is fixed: lens planes including an unusually large number of massive halos will have room for fewer low-mass halos, and vice-versa. Also, as we discuss in Sec. 3.4, the covariance underestimation can be the consequence of halos being randomly placed in the field of view.

To analyze the cosmology dependence of the covariance, we plotted the value of selected normalized matrix elements as a function of  $\Sigma_8$  for all cosmologies in Fig. 3.4. For N-body data, all variances and correlations increase until  $\Sigma_8 \approx 0.6$  and then plateau. This dependence may affect the likelihood calculations. Matrices computed with CAMELUS show a very weak cosmology dependence and all their elements are smaller –in absolute value– than those from N-body simulations, which would result in lower error estimations.

After comparing peak counts and their covariances, we combined these to estimate the  $L_{cg}$  likelihood for each model. We show the  $2\sigma$  (95.4%) credible contours in Fig. 3.5 by numerically integrating the interpolated likelihoods, and compared the results in Table 3.4. We find thicker contours, with a 30% larger overall area, which can be attributed to the larger covariances (see below).

We also report any shifts in the credibility region’s centroid position in Table 3.4. The centroid is defined as the point whose position is the arithmetic mean of that of all points within the region:

$$\theta^{centroid} = \frac{\int_{CR} d\theta d\Theta \theta}{Area_{CR}} \approx \frac{\sum_{CR} \theta^i}{\sum_{CR} 1} \quad (3.18)$$

where  $\theta$  refers to the axis for which the centroid coordinate is computed and  $\Theta$  to all other dimensions in parameter space. We did not use the maximum likelihood to estimate shifts because it corresponds to the fiducial cosmology by construction. We found a significant shift exclusively between N-body contours computed using all the peaks and those computed using only high-significance peaks. The contours from N-body simulations are more tilted in  $\{\Omega_m, \sigma_8\}$ . To quantify the difference in tilt, we fitted the exponent ( $\alpha$ ) of the degeneracy relation,  $\Sigma_8 \equiv \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^\alpha$  to

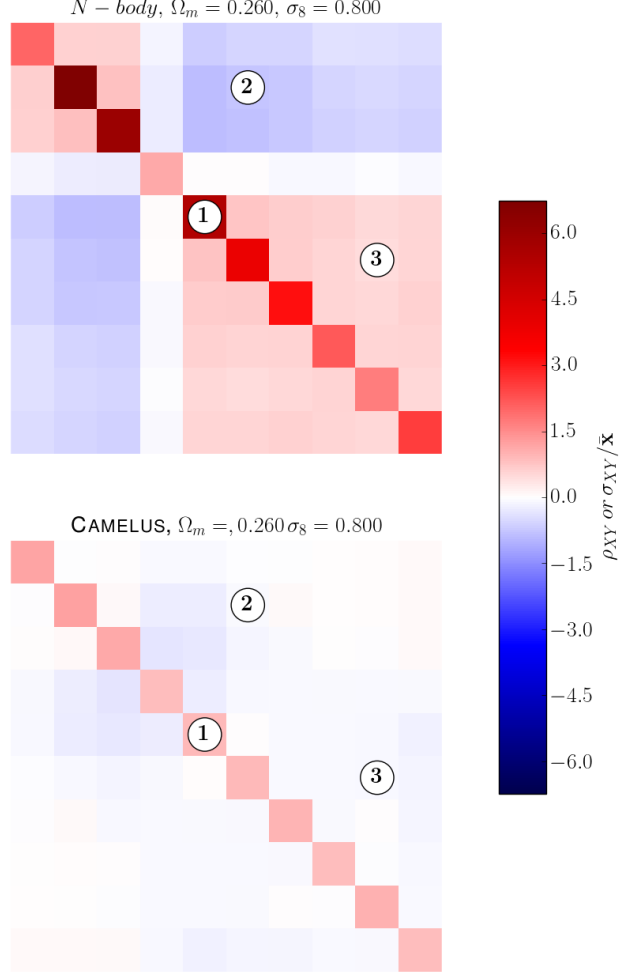


Figure 3.3: Covariance comparison between N-body (upper panel) and CAMELUS (lower panel) for the fiducial cosmology. Each normalized covariance matrix has diagonal elements equal to the peak count variance divided by its mean,  $\frac{\sigma_{ii}^2}{\bar{x}_{ii}}$ , and off-diagonal elements equal to the correlation coefficients,  $\rho_{ij} \equiv \frac{\sigma_{ij}}{\sigma_i \sigma_j}$ . We find higher absolute values for all elements in the matrices, with positive and negative peaks positively correlated and positive peaks anti-correlated with negative ones. Peak counts from CAMELUS are mildly anti-correlated. Selected matrix elements whose value for all cosmologies is displayed in Fig. 3.4 are indicated with a number.

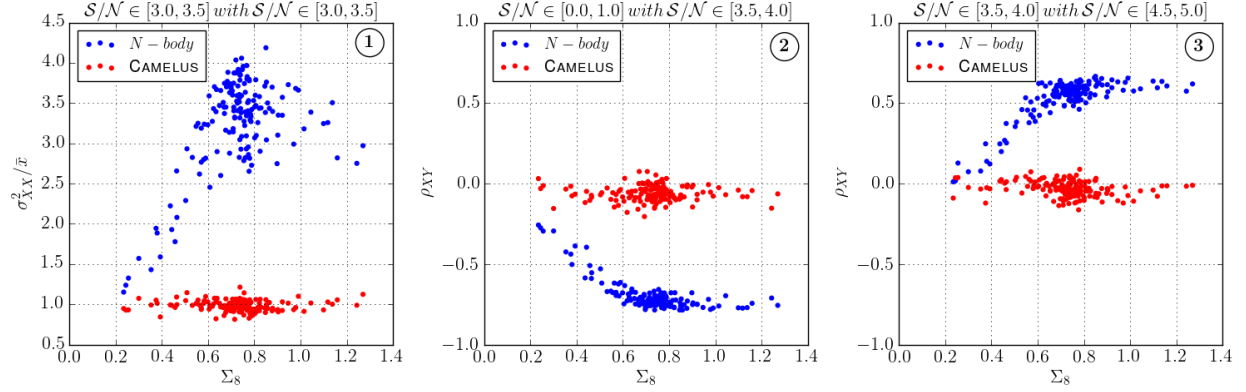


Figure 3.4: The cosmology dependence of covariances. Each subplot shows the value of selected normalized covariance matrix elements for all 162 cosmologies. The selected elements are indicated in Fig. 3.3, and correspond to a diagonal element (left panel) and off-diagonal elements showing anti-correlation in N-body data (center panel) and correlation (right panel). N-body data exhibit higher absolute values for all elements and stronger cosmology dependence.

minimize the scatter in  $L_{cg}$ . We restricted the data to  $\Sigma_8 \in [0.6, 0.9]$ , since estimating the scatter for extreme values of  $\Sigma_8$  where we have few data points is problematic. We find an exponent of  $\alpha = 0.67$  vs.  $\alpha = 0.58$  for CAMELUS.

It is common to restrict analyses to the highly significant peaks, since their counts are not dominated by shape noise. We emphasize that the shape noise can be measured accurately from the data themselves, and so there is no reason a priori to discard the 'noisy' peaks with a lower  $S/N$ . Nevertheless, we investigated the impact of this restriction. We find that it does not change the contours obtained with CAMELUS, but has a drastic impact on those from N-body simulations, as can be seen in Fig. 3.5. Previous works [31, 32] found that low-significance peaks carry important cosmological information in WL maps from N-body simulations. Table 3.4 shows that the contours double in size when only peaks with  $S/N > 3$  are considered. While both models yield similar constraints, they derive their predictive power from different  $S/N$  peaks.

Finally, we assessed the impact of using a variable covariance matrix when computing the likelihood in the same way as was done in [154]. Estimating the covariance at each point of the parameter space is computationally expensive, but as we have shown, the covariance can change significantly. Fig. 3.6 shows the effect on both  $1\sigma$  (68.3%) and  $2\sigma$  (95.4%) contours; the values for

	$FoM$	$\Delta Area$	$\Delta\Omega_m$	$\Delta\sigma_8$
$N - body\ all\ peaks$	26	-	-	-
$N - body\ S/N > 3.0$	9	+198%	+0.05	-0.09
CAMELUS all peaks	36	-28%	+0.02	-0.00
CAMELUS $S/N > 3.0$	33	-21%	+0.03	-0.02

Table 3.4: Comparison of  $L_{cg}$   $2\sigma$  (95.4%) credible contours. The figure-of-merit,  $FoM$ , is the inverse of the area of the credibility regions. Also displayed are the percentage changes in the area of the credibility region and its centroid (arithmetic mean) shift. We find looser constraints ( $\approx 30\%$ ) for N-body data, whose predictive power is greatly diminished when low significance peaks are excluded from the analysis.

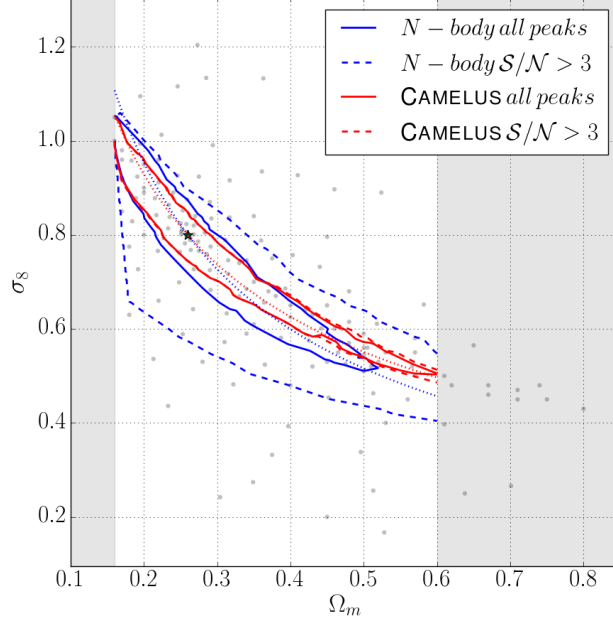


Figure 3.5: Comparison of  $2\sigma$  (95.4%) credible contours from N-body (blue) and CAMELUS (red) data, using a Gaussian likelihood with constant covariance,  $L_{cg}$ . Solid lines show the contours computed using all the peak counts. We find looser constraints, with a thicker,  $\approx 30\%$  larger credibility region. Dashed lines show the results including only high significance peaks ( $S/N > 3$ ). While constraints based on CAMELUS data do not change, the predictive power from N-body data is severely reduced, with a  $\approx 200\%$  increase in the area of the credibility region. Dotted lines show the degeneracies  $\Sigma_8 = \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^\alpha$  that minimize scatter in  $L_{cg}$ . We find a steeper contour,  $\alpha = 0.67$  vs.  $\alpha = 0.58$  for CAMELUS. Grey dots show the simulated cosmologies (a green star the fiducial cosmology), and grey areas the regions excluded from contour measurements.

$\mathcal{S}/\mathcal{N} > 3$ Peaks				
Likelihood	FoM	$\Delta Area$	$\Delta \Omega_m$	$\Delta \sigma_8$
$1\sigma L_{cg}$	25	-	-	-
$1\sigma L_{svg}$	29	-14%	+0.01	+0.02
$1\sigma L_{vg}$	36	-19%	-0.03	+0.01
$2\sigma L_{cg}$	9	-	-	-
$2\sigma L_{svg}$	13	-32%	-0.01	+0.09
$2\sigma L_{vg}$	15	-13%	-0.01	-0.01
All peaks				
Likelihood	FoM	$\Delta Area$	$\Delta \Omega_m$	$\Delta \sigma_8$
$1\sigma L_{cg}$	69	-	-	-
$1\sigma L_{svg}$	81	-14%	-0.00	+0.01
$1\sigma L_{vg}$	81	-0%	+0.00	-0.01
$2\sigma L_{cg}$	26	-	-	-
$2\sigma L_{svg}$	32	-19%	-0.02	+0.05
$2\sigma L_{vg}$	32	+1%	+0.00	-0.01

Table 3.5: Effect of using a cosmology-dependent covariance matrix.  $1\sigma$  (68.3%) and  $2\sigma$  (95.4%) credible contours are computed using the three likelihoods described in § 3.2.3 ( $L_{cg}$ ,  $L_{svg}$  and  $L_{vg}$ ). The analysis is done twice, using only high significance peaks ( $\mathcal{S}/\mathcal{N} > 3$ ) and all the peaks. We report the figure of merit (FoM); defined as the inverse of the area of the credibility region), changes in the credibility regions and shifts in their centroid. Introducing a cosmology-dependent covariance into the  $\chi^2$  term of the Gaussian likelihood has a bigger impact than introducing it in the determinant term. Also, the effect is bigger when only high peaks are included.

the changes are listed in Table 3.5. The effects are always more important if only high-significance peaks are included. Introducing a variable covariance in the  $\chi^2$  term of a Gaussian likelihood – i.e., using  $L_{svg}$  instead of  $L_{cg}$ – tightens constraints by 14 – 19% (14 – 32% for high  $\mathcal{S}/\mathcal{N}$  peaks only). Incorporating it also to the determinant term –i.e., going from  $L_{svg}$  to  $L_{vg}$ – has a more limited impact of 0 – 1% (13 – 19% for high  $\mathcal{S}/\mathcal{N}$  peaks-only). It would be advisable then to use a cosmology-dependent covariance for a precise determination of parameter constraints, with the exception of those cases in which most of the parameter space has been rejected by previous experiments and only a small region needs to be explored.

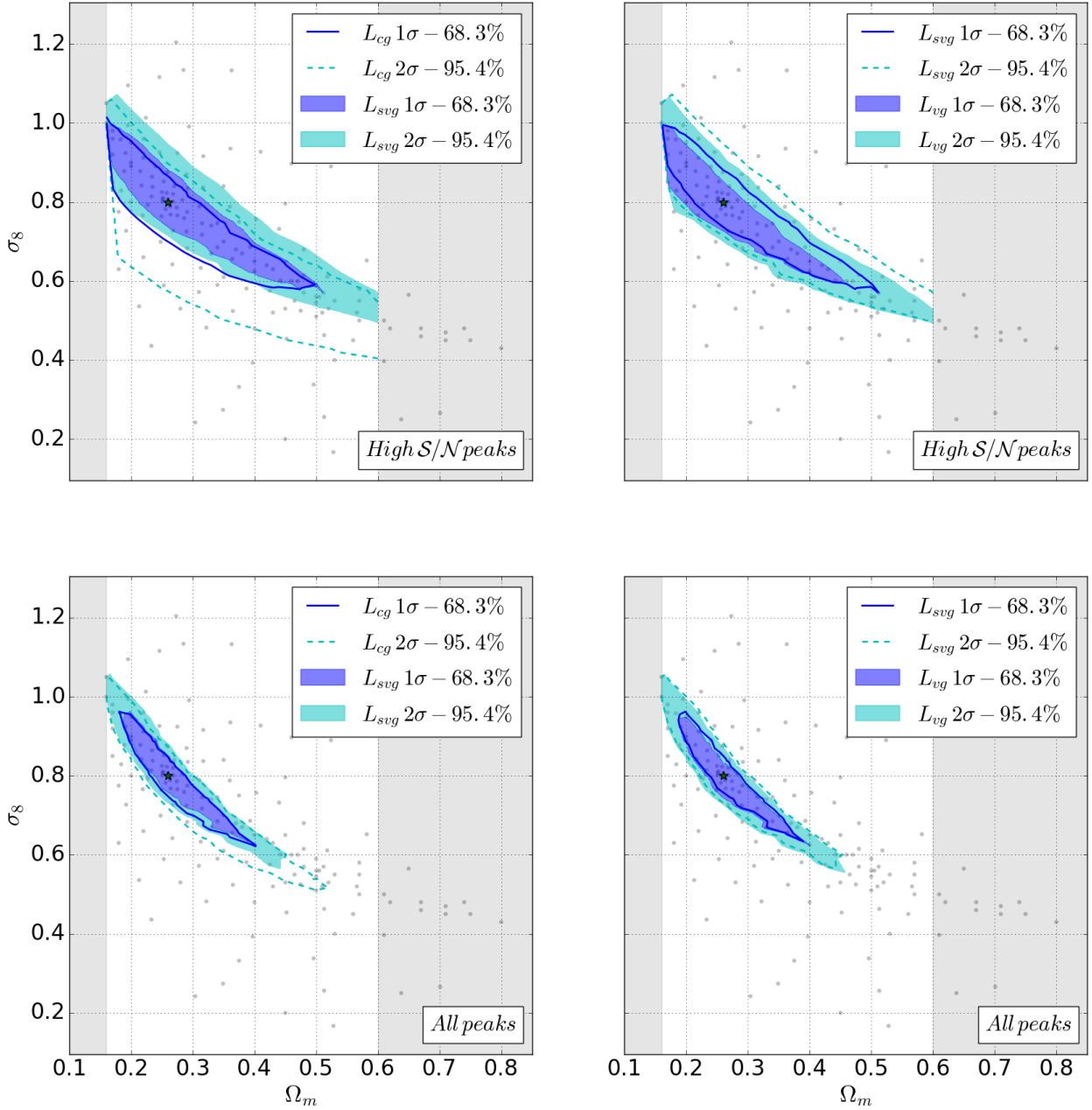


Figure 3.6: Effect on the credibility regions of using a cosmology-dependent covariance. In the left panels we show the change caused by introducing a variable covariance in the  $\chi^2$  term of a Gaussian likelihood ( $L_{svg}$ , shaded areas) compared with a constant covariance ( $L_{cg}$ , lines). On the right we display the change from using a variable covariance matrix in the determinant term as well ( $L_{vg}$ , shaded areas) compared with  $L_{svg}$  (lines). The upper panels show the result using only high-significance ( $S/N > 3$ ) peaks, while the lower panels show results with all peaks included. Introducing a variable covariance in the  $\chi^2$  has a larger impact than using it in the determinant term. Also, the effects are larger when using only high significance peaks (see Table 3.5).

### 3.4 Discussion

Given the restricted scope of this paper –to assess the accuracy of the halo-based model CAMELUS for cosmological inference using WL peaks– our main findings are the differences between its credible contours and those from N-body simulations.

We identified small discrepancies in peak counts and significantly larger covariances from N-body data, with a stronger dependence on cosmology. To disentangle the effect of both elements on parameter inference, we computed "hybrid" likelihoods mixing peak counts from one model with covariance matrices from the other. Fig. 3.7 shows the resulting  $2\sigma$  credibility regions. Substituting the covariance for that from CAMELUS data shrinks the N-body contours to a thickness equivalent to that of CAMELUS. The effect on the credibility region from CAMELUS of using peak counts from N-body simulations is comparatively less important, suggesting that more accurate estimation of covariances have the highest potential for improvement. The upper panels of Fig. 3.7 were plotted using only high significance peaks and show even more clearly how differences in the covariance matrices drive the size and shape of the credible contours.

To understand the origin of the discrepancy in peak-count variance, we compared halo counts from both models, since there is an established connection between halos and convergence peaks [160]. To identify halos in our N-body simulation we used the Amiga Halo Finder (AHF) [175]. Since we evolved a single  $240 h^{-1}\text{Mpc}$  box per cosmology, we subdivided it into sub-volumes to compute the variance. We split our simulation volume in  $3^3$ ,  $4^3$  and  $5^3$  equally sized sub-boxes and scaled the counts to a common reference volume. We ran CAMELUS to generate halo catalogues corresponding to similar volumes as those of the sub-boxes used for the N-body calculation, and scaled the counts in the same way. The results are shown in Fig. 3.8 and are in good agreement with analogous findings for cluster counts [176]. Cumulative halo counts from N-body simulations have a higher sample (cosmic) variance than what would be expected if it were due solely to shot noise that follows a Poisson distribution. We use [172] for the mean counts in the shot noise calculation. This is the same halo mass function used in CAMELUS, and we verified that it



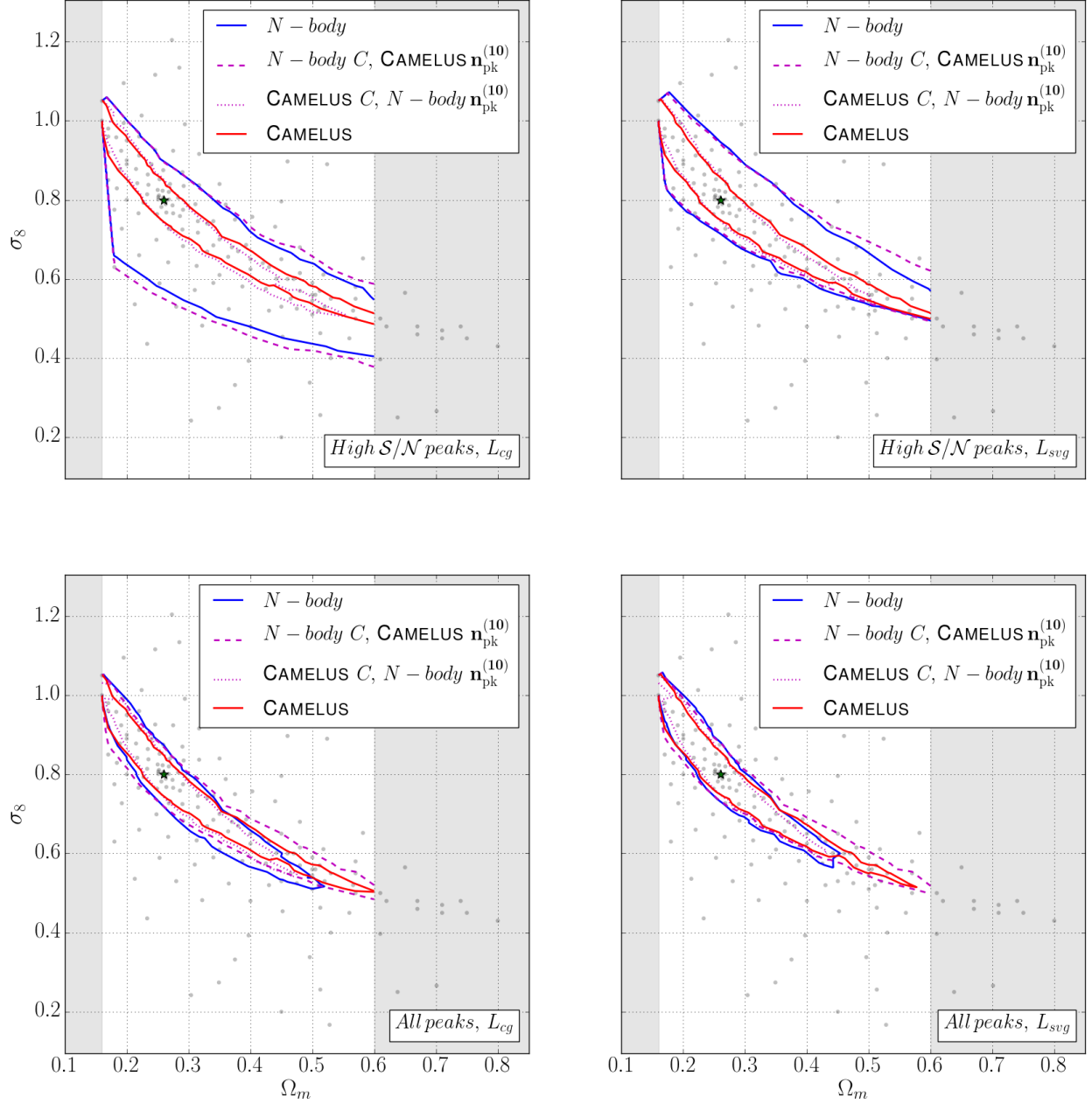


Figure 3.7: Impact of differences in peak counts and covariance matrices on credible contours. Solid lines are  $2\sigma$  contours from N-body (blue) and CAMELUS (red) data. Magenta lines are contours computed mixing peak counts from one model with the covariance matrices from the other. The dashed contours combine N-body covariance matrices with CAMELUS peak counts, and the dotted contours combine conversely N-body peak counts with CAMELUS covariances. The upper panels show the results using only  $S/N > 3$  peaks while the lower panels display the contours obtained including all peaks. On the left we show contours computed using a constant covariance,  $L_{cg}$  and on the right those introducing a variable covariance in the  $\chi^2$  term,  $L_{svg}$ . In general, contours computed with the same covariance matrices are closer than those calculated with the same peak counts. The effect is more noticeable for the cases which include only high-significance peaks, since for these the N-body and CAMELUS contours exhibit a greater difference.

was in good agreement with the halos extracted from our N-body simulation. The excess sample variance is caused by LSS clustering halos which increases the correlation of their positions. As halos become more massive and rarer, shot noise becomes more important and the excess sample variance diminishes.

CAMELUS places halos randomly, and its halo sample variance is dominated by shot noise except for the low-mass tail of the halo distribution. Halos are sampled from an analytical mass function until the total mass in a volume reaches its expected mean value. This condition that the total mass in halos is fixed links high- and low-mass halo numbers, transferring variance to the low-mass halo range. Nevertheless, this effect does not translate into larger covariances, since low-mass halos do not contribute to peak counts. We compared peak counts from CAMELUS using different minimum halo masses ( $10^{10}$ ,  $10^{11}$  and  $10^{12} M_{\odot}$ ) and found virtually no difference.

Convergence peaks resulting from the projected mass density field, exhibit a similar pattern. The upper panel of Fig. 3.8 shows the variance in the cumulative peak counts as a function of their height. Peak counts from N-body data also have a higher sample variance compared to a Poisson distribution and, as the peak  $S/N$  increases, shot noise becomes more important. For CAMELUS data, sample variance is smaller and is dominated by shot noise. The counts come from 500  $3.5 \times 3.5 \text{ deg}^2$  convergence maps for the fiducial cosmology.

The parallel between halo and peak-count sample variance suggests that modifying the CAMELUS algorithm to account for halo clustering could enhance its accuracy by yielding larger covariance matrices that would propagate into looser parameter constraints.

We also found that including low-significance peaks in the analysis improves the predictive power for N-body simulations, while it does little for CAMELUS. Fig. 3.9 and Table 3.6 show the effect of adding bins of decreasing significance peaks to the contours' computation. For N-body simulations, the impact is particularly important when peaks in the range  $S/N \in [2, 3]$  are incorporated, with  $2\sigma$  contours reduced by 25 – 48%. Those moderately low-significance peaks have been associated with constellations of small halos [159, 160]. These alignments are missing in the halo catalogs generated with CAMELUS, which constrains cosmology essentially through

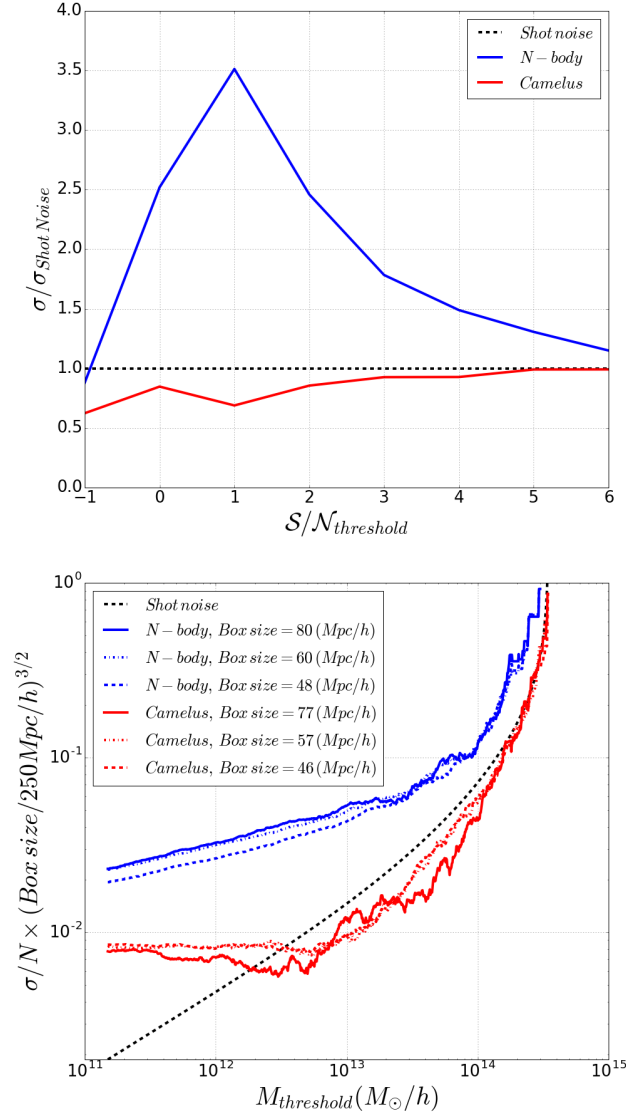


Figure 3.8: Peak and halo count variance comparison between N-body (blue) and CAMELUS (red). **Upper panel:** ratio of the cumulative peak count standard deviation from its value expected for a Poisson distribution, as a function of peak height. For pure Poisson shot noise, this ratio is unity (horizontal black dashed line). We find significantly higher sample variance than the results from CAMELUS, and what would be expected for a Poisson distribution. As the peak height increases and the peak counts decrease, shot noise starts to dominate. **Lower panel:** variance of the cumulative halo number as a function of minimum halo mass. Sample variance is estimated from different sub-volumes, and scaled to a common reference volume of  $(250 h^{-1} \text{Mpc})^3$ . Shot noise is estimated from a Poisson distribution with mean value adopted from a theoretical halo mass function [172]. N-body cumulative halo counts exhibit a sample variance higher than expected from a Poisson distribution. Shot noise becomes more important at higher masses, as the halos become scarcer. CAMELUS is dominated by shot noise.

$N - body$	$L_{cg}$		$L_{svg}$		$L_{vg}$	
	FoM	$\Delta(\%)$	FoM	$\Delta(\%)$	FoM	$\Delta(\%)$
$S/N > 4.0$	9	-	11	-	13	-
$S/N > 3.5$	8	-0	12	-2	14	-4
$S/N > 3.0$	9	-4	13	-11	15	-9
$S/N > 2.0$	17	-48	19	-32	20	-25
$S/N > 1.0$	22	-22	27	-31	27	-27
$S/N > 0.0$	24	-10	31	-11	30	-10
$S/N > -1.0$	25	-3	31	-1	31	-1
$S/N > -\inf$	26	-6	32	-3	32	-4

CAMELUS	$L_{cg}$		$L_{svg}$		$L_{vg}$	
	FoM	$\Delta(\%)$	FoM	$\Delta(\%)$	FoM	$\Delta(\%)$
$S/N > 4.0$	26	-	27	-	27	-
$S/N > 3.5$	30	-12	30	-10	30	-10
$S/N > 3.0$	33	-10	33	-9	34	-9
$S/N > 2.0$	35	-4	36	-8	36	-8
$S/N > 1.0$	36	-3	37	-2	37	-2
$S/N > 0.0$	36	-1	37	-1	37	-1
$S/N > -1.0$	36	+0	37	-0	38	-0
$S/N > -\inf$	36	-0	38	-0	38	-1

Table 3.6: Impact on the models’ predictive power of the lowest significance peak bin included in analysis. Figure of merit (FoM) and change in  $2\sigma$  contour area ( $\Delta\%$ ) for constant, semi-varying and variable covariance likelihoods.

high peaks which are caused by high-mass halos.

Our likelihood calculations rely on a precise estimation of the precision matrix,  $C^{-1}$ , and the determination of the credible contours on the interpolation of the likelihood beyond the discrete set of cosmologies for which we run simulations.

For each cosmology, we estimated the covariance matrices using 500 converge field realizations recycled from a single N-body calculation by slicing, shifting and rotating the simulated box. Previous work showed [145] that a single N-body run is sufficient to generate  $\approx 10^4$  convergence maps whose peak counts are statistically independent, and two boxes would be enough to measure feature means with an accuracy of 50% of the statistical error. Therefore, we decided to use a single box, which allowed us to maximize the number of cosmologies to sample given our available computing resources.

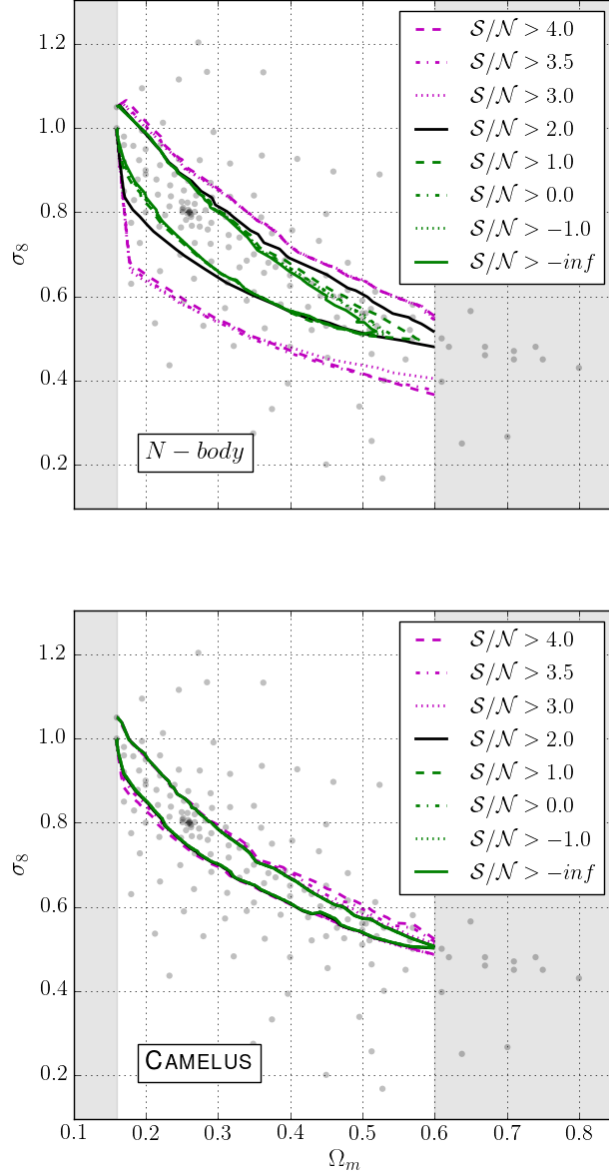


Figure 3.9: Influence on the credibility region of the lowest significant peaks included in the ( $L_{cg}$ ) likelihood calculation. **Upper panel:** for N-body simulations, including peaks with  $2.0 < \mathcal{S}/\mathcal{N} < 3.0$  significantly improves the model's predictive power. **Lower panel:** for CAMELUS, little or no improvement in predictive power is found when lower-significance peaks are included.

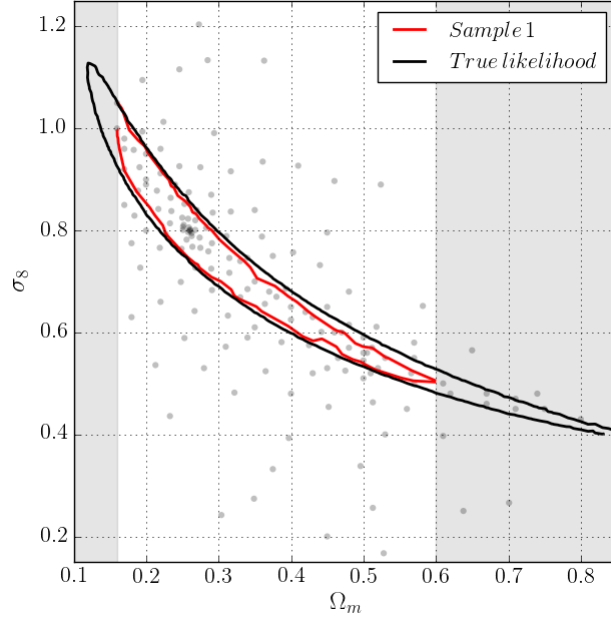


Figure 3.10: Effect on CAMELUS credible contours of finite sampling of the cosmological parameter space.  $2\sigma$  contours obtained from a fine grid of 7,803 models (black) and interpolated from our suite of 162 cosmologies (red). The interpolated contour is smaller in the low- and high- $\Omega_m$  tails. Thus we excluded from our analyses the greyed-out regions, corresponding to  $\Omega_m < 0.160$  and  $\Omega_m > 0.600$ .

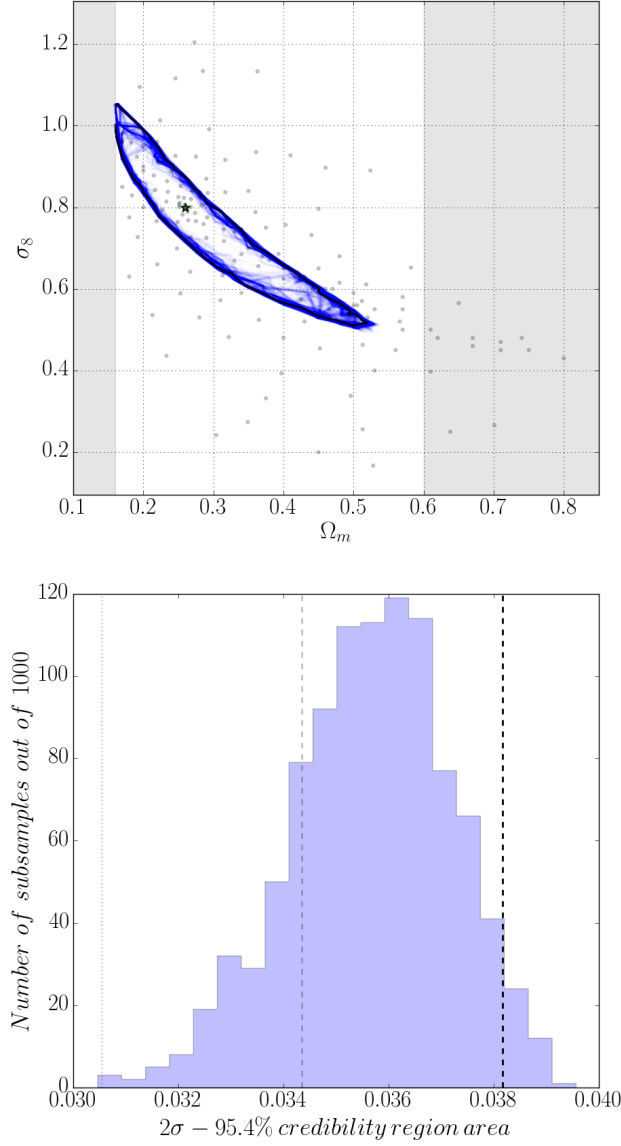


Figure 3.11: Effect of cosmological parameter sampling on the N-body credible contours. We draw 1,000 bootstrap samples from our suite of cosmologies; i.e., we draw samples of 162 elements with substitution, each having on average 102-103 different cosmologies. **Upper panel:**  $L_{cg}$   $2\sigma$  contours from the full suite (black) and the 1,000 sub-samples (blue). Darker areas indicate higher contour concentration. **Lower panel:** area histogram for the bootstrap samples. Displayed for reference are the area for the full suite (black dashed line), 90% of this value (grey dashed line) and 80% (grey dotted line). 81% of the contours fall within 10% of the original area and 99% within 20%.

While the inverse of a covariance matrix estimated from data is not unbiased, since the number of realizations we use (500) is much larger than the dimension of our data vectors (10), the bias is negligible ( $\approx 2\%$ ). We verified that the results with Gaussian likelihoods after de-biasing the covariances following [151] were the same as those from using the non-Gaussian form of the likelihood found in [174].

Interpolation can also introduce errors in the contours. We verified this effect on the CAMELUS contours by running an additional fine grid of 7,803 cosmologies –described in Fig. 1 of ref. [154]–, and plotting the contours obtained from these and our original models in Fig. 3.10. The regions corresponding to low- and high- $\Omega_m$  values are under-sampled, and as a result the contours in those regions are underestimated. Therefore, we limited our contour analyses to the interval  $\Omega_m \in [0.160, 0.600]$ , where the true and the estimated contours agree within 20%.

Since we could not reproduce this analysis for our N-body simulations due to the computational cost, we generated contours from bootstrap samples of our full simulation set. That is, we sampled from the 162 cosmologies, with substitution, and drew the resulting contours in Fig. 3.11. Each sample had an average of 102-103 unique cosmologies in them. As with the analysis of the effect of sampling on the CAMELUS contour, we found that dropping models almost always results in a smaller area, and as a result we may be underestimating the errors on the parameters. We expect that underestimation to be moderate, for 81% of the samples yield areas that lie within 10% of the area computed with the full sample and 99% of the samples fall within 20%. The highest risk is missing part of the tail of the credibility region, which occurs in some of the random bootstrap samplings.

We do not address the question of whether a Gaussian likelihood is an appropriate model for our data, since the focus of this study is to compare the results from the two models. We will treat it in future work. For CAMELUS data, the Gaussian approximation yields credible contours in good agreement with those computed using the actual distribution of peak counts, as can be seen in the left panel of Fig. 8 in [154].

Other underlying simplifications common to both the N-body and halo-based simulations used



in this work are the non-inclusion of baryonic effects, the Born and flat sky approximations, and the omission of any survey effects such as masking, instrument systematics, etc. Baryons have been shown to increase the amplitude of the WL power spectrum on small scales and to introduce a small bias in high  $S/N$  peaks [177]. The precision requirements and large sky coverage from future surveys will require the inclusion of these baryonic effects [178], as well as revisiting some of the approximations used in our models [5].

In future work, new modified ray-tracing simulations using manipulated snapshots from N-body simulations may clarify the specific sources of discrepancy between N-body and halo-based models. Possible reasons can be enumerated as follows:

- (i) Non-halo contributions, e.g. filaments, walls,
- (ii) Halo clustering,
- (iii) Non NFW halo profiles, e.g. merging halos, triaxiality, and
- (iv) Halo concentration; e.g., broad distribution instead of a deterministic function.

Modifications to a model such as CAMELUS to address points (ii)-(iv) could in principle be addressed within the halo model framework and would make it even more useful as a fast lensing emulator by improving its accuracy.

### 3.5 Conclusions

In this work we compared the outcomes from the fast halo-based algorithm CAMELUS with those of N-body simulations for a suite of cosmologies spanning a wide range of values in the  $\{\Omega_m, \sigma_8\}$  plane.

We found larger (by  $\approx 30\%$  in area), more significantly tilted (by  $\approx 13\%$  in angle) credible contours from N-body data. Importantly, the two models draw their predictive power from a different types of peaks. While CAMELUS constrains cosmology through high- $S/N$  peaks associated with massive halos, the N-body data are highly sensitive to lower- $S/N$  peaks.

The larger thickness and overall area of the N-body credible contours are mostly driven by the covariances, with peak counts showing a higher variance than expected from pure shot noise. This suggests that modifying the placement of halos in CAMELUS to account for the correlations in their locations is a promising way to improve its covariance estimation and accuracy as a WL peak count emulator.

Using a cosmology-dependent covariance matrix for likelihood estimation improves constraints by 14 – 20%, and thus will be needed in order to achieve high-precision parameter estimations.

Finally, we have found that optimal sampling of a high-dimensional parameter space with expensive N-body simulations to define credibility regions with high precision is a topic that requires further investigation, and a fast simulator like CAMELUS could prove itself particularly valuable by providing a first estimation of the likelihood.

## Chapter 4: Optimizing simulation parameters for weak lensing analyses involving non-Gaussian observables

### 4.1 Introduction

Weak gravitational lensing (WL) enables the mapping of the distribution of dark matter (DM) in the universe on large scales and as a result is a powerful probe to infer cosmological parameters such as  $\sigma_8$  and  $\Omega_m$  (see comprehensive reviews by, e.g. [4], [179] and [5]). In practice, the lensing signal can be extracted from statistical measurements of the shapes of background galaxies, distorted by deflections in the path of light rays as they traverse the vicinity of matter over- and under-densities.

Upcoming surveys such as DESI [180], LSST [21], Euclid [22], WFIRST-AFTA [24] and SKA [181], will provide WL data of unprecedented quality and quantity and will require correspondingly accurate and precise models to extract information from these datasets. WL observables delve into the non-linear regime, which can be forward-modeled by ray-tracing photons through high-resolution simulated dark matter (DM) density fields [182, 183, 184, 185, 186]. Simulating the large volumes encompassed by future surveys is computationally expensive, especially when a high-dimensional parameter space needs to be explored. Different ideas to reduce the cost of forward modeling WL observables have been put forward and tested. These include using approximate codes to simulate the evolution of the matter density field – for example ICECOLA [187], L-PICOLA [188] or FastPM [189]–, analytic or semi-analytic models – for example CAMELUS [170] and machine learning approaches – for example generative adversarial networks (GANs; e.g. [190, 110, 111]). While analytic models can predict two-point statistics with sufficient accuracy [191], higher-order statistics which capture non-Gaussian information from the non-linear regime require a numerical approach [192, 42, 142].

In this paper we study, within the framework of raytracing N-body simulations, the sensitivity of the power spectrum (PS) and several non-Gaussian statistics commonly used in WL studies: the one-point probability density function (PDF), lensing peak counts, and the full set of Minkowski functionals (MFs) to the two hyper-parameters with the highest impact on the computational cost of the simulations: (i) the thickness of the lens planes used to construct the past light cones in raytracing and (ii) the mass resolution of the N-body simulations used to model the underlying matter density field. Previous studies have already tackled some of these or related aspects. For instance, [183] verified the effect on the measured PS of the mass resolution of the underlying DM simulation, the grid size and interpolation method used when raytracing and also studied the contribution of super-sample modes to the PS variance. [192] looked at the effect of the resolution of the 2D lens planes on the measured convergence power spectrum, and evaluated the non-Gaussian contribution to its covariance matrix. [184] investigated the effect of mass resolution and comoving distance between lens planes on the convergence power spectrum, skewness and kurtosis. This work revisits the sensitivity of the measured convergence power spectrum to the mass resolution and distance between lens planes, and extends the analysis of numerical convergence analyses to non-Gaussian statistics within the framework of a tomographic WL analysis of a LSST-like survey.

The manuscript is organized as follows. In § 4.2 we describe our simulation pipeline (§ 4.2.1), how we quantify the impact of the hyper-parameters (§ 4.2.2), and the statistics used to assess that impact (§ 4.2.4). In § 4.3 we report and discuss the results of the analysis. Finally, we summarize our conclusions in § 4.4.

## 4.2 Methods

### 4.2.1 Simulating convergence maps

Our analysis is based on lensing statistics measured on convergence maps obtained from raytracing dark matter only N-body simulations. Since weak lensing probes large-scale structures in the non-linear regime, direct simulations offer a way to characterize the WL signal.

The N-body simulations are run using the publicly available Tree-PM code *Gadget2* [144],

which evolves Gaussian initial conditions generated with `NGenIC` [193]. The initial conditions are defined by power spectra computed with `CAMB` [194]. The positions of the particles at different redshifts are used to build past light-cones. The trajectory of a bundle of rays is followed along past light-cones according to the multi-plane algorithm to generate synthetic convergence maps. We refer the reader to [143] for a detailed step-by-step description of our pipeline and its implementation in `LensTools`, and to [183] for a review of the theoretical basis of the algorithms used.

All the simulations have the same comoving volume of  $(240h^{-1} \text{ Mpc})^3$ , phases for the initial conditions, and underlying cosmology. The cosmological parameters are consistent with [127]:  $(H_0, \Omega_m, \Omega_\Lambda, \Omega_b, w_0, \sigma_8, n_s) = (67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}, 0.309, 0.691, 0.0486, -1.0, 0.816, 0.967)$ . As an illustration, we compare the convergence power spectrum measured in our highest-resolution simulation ( $1024^3$  DM-only particles, or a particle mass of  $1.1 \times 10^{10} M_\odot$ ), with flat-sky approximation predictions (see Eq. (4.3)) computed using the `halofit` model [146, 195], as implemented in the Core Cosmology Library (CCL; [196]). In data smoothed at 1 arcmin resolution with the presence of the shape noise levels considered here, the agreement between theory and simulations stays within 20% for  $\ell < 12,000$ . In noiseless data, a loss of power in the simulations is noticeable, especially for sources at low redshift. The worst case corresponds to the bin with sources at  $z = 0.5$ , which shows a 50% loss in power at  $\ell \approx 7,500$ . In all cases we measure excess power at large angular scales relative to the theoretical expectation (up to  $\approx 20\%$  for the noiseless case and  $\approx 13\%$  in the presence of shape noise). The reason of this excess is the sample variance in the matter power spectrum, as low multipoles probe scales that are comparable with the size of our simulated volume. Since the purpose of this study is the calibration of hyper-parameters, we are interested in relative changes and the sample variance at large scales should not impact our conclusions. Furthermore, we are mostly interested in relative changes on small scales, where the use of non-Gaussian observables is relevant. To perform inference from data, this sample variance would need to be tackled, for instance, by simulating a larger volume (or many independent N-body realizations of our simulation volume).

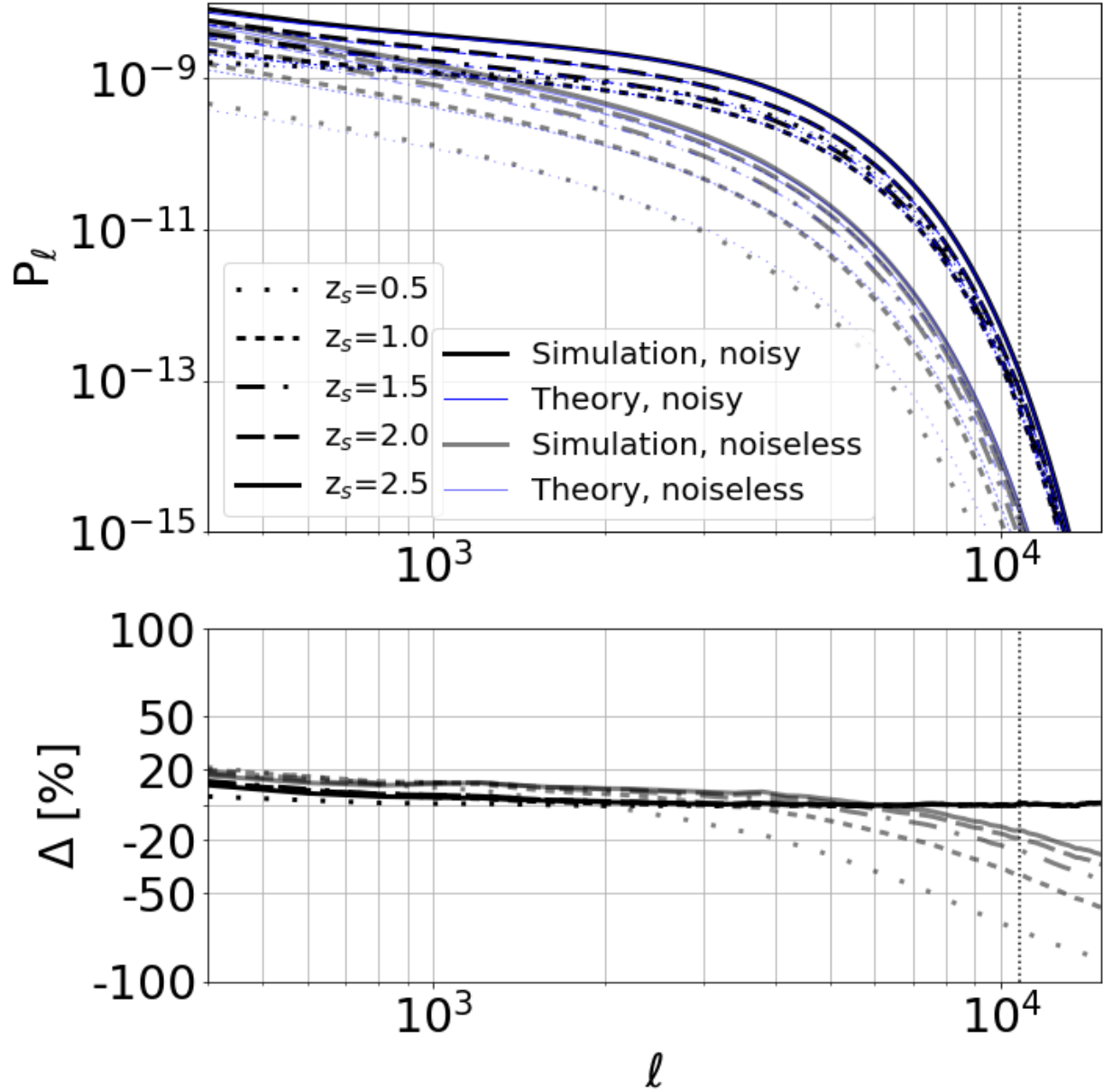


Figure 4.1: Comparison of convergence power spectra measured in our simulations with  $1024^3$  DM particles, to those obtained in semi-analytic calculations using the flat-sky approximation and `halofit`. **Upper panel:** Convergence power spectra. Thick, black lines correspond to simulations, while thin, blue lines correspond to the `halofit` calculations. Different dashes indicate different redshift bins for the lensed galaxies. Full lines correspond to results in the presence of shape noise (see section § 4.2.1 for a description of the shape noise level considered), and partly transparent lines to the results in the absence of noise. **Lower panel:** percentage difference between simulations and `halofit`, for each redshift bin. As in the upper panel, full lines correspond to the noisy, and partly transparent lines to the idealized noiseless case.

For each configuration, a single volume is simulated and reused through random shifts and rotations to generate as many as  $O(10^4)$  pseudo-independent past light-cones [145]. Since each convergence map covers only  $3.5 \times 3.5 \text{ deg}^2$  on the sky at this redshift, the flat-sky approximation holds. To account for their intrinsic ellipticity, Gaussian random shape noise is added independently at each of the  $1024 \times 1024$  pixels in each convergence map. The standard deviation of this noise,

$$\sigma_{\text{pix}} = \sqrt{\frac{\sigma_\epsilon^2}{2n_g A_{\text{pix}}}}, \quad (4.1)$$

depends on the variance of the galaxies' intrinsic ellipticity (assumed to be  $\sigma_\epsilon = 0.35$ ), the solid area covered by a pixel ( $A_{\text{pix}} = 0.04 \text{ arcmin}^2$ ), and the effective galaxy number density.

We consider an LSST-like survey with a footprint of  $20,000 \text{ deg}^2$  and galaxy density distribution  $n(z) \propto z^2 \exp[-2z]$  [21]. Because a small number of redshift bins suffices to extract most of the tomographic information encoded in weak lensing data sets [155], we use five tomographic bins. In each bin, we assume all lensed galaxies are at a fixed redshift of  $[0.5, 1.0, 1.5, 2.0, 2.5]$ . The galaxy densities we consider in each redshift bin are  $[8.83, 13.25, 11.15, 7.36, 4.26] \text{ arcmin}^{-2}$ .

A final smoothing is applied to both noiseless and noisy maps with a Gaussian kernel of 1 arcmin standard deviation. While only results from maps with shape noise are relevant for the analysis of future survey data, we show also results for smoothed, noiseless convergence maps, since the effect of different simulation choices are often more discernible in those.

#### 4.2.2 Assessing the impact of hyper-parameters

The accuracy of the forward model as a function of different values of the hyper-parameters is assessed by comparing the statistics of observables measured over 10,048 convergence maps simulated for each configuration. As explained in § 4.2.1, for each set of hyper-parameters, all 10,048 maps are generated from a single, recycled, N-body simulation.

We chose a "fiducial" configuration as a reference. The difference between an observable's mean for all cases and the fiducial model is compared with a standard error. As standard error, we consider 3 standard deviations measured on the fiducial model's maps, scaled to a survey sky area

of  $2 \times 10^4 \text{ deg}^2$  (commensurate with LSST's). This scaled standard deviation represents a lower bound on the uncertainty expected in future surveys, as it includes only the statistical error from intrinsic ellipticities. For a review of all sources of error in WL surveys, see for instance [49].

Ultimately, the most relevant metric is how much the lens plane thickness affects the inference of cosmological parameters. A definite answer to this question requires the calculation of the credible contours for the parameters of interest. Within the scope of our single-cosmology numerical experiments, we instead look at the reduced  $\chi^2$ , which combines differences in the mean observables with their covariance matrix.

In general, a given observable is a vector  $\mathbf{s}$  with components corresponding to bins of spherical harmonic index  $\ell$  for the power spectrum, or  $\kappa$  thresholds for the other statistics, such as the one-point PDF, lensing peaks or Minkowski functionals. We compute the  $\chi^2$  statistic for each configuration, considering the mean of the fiducial model as ground truth:

$$\chi^2 = (\mathbf{s} - \mathbf{s}_{fid})^T \widehat{\mathbf{C}}^{-1} (\mathbf{s} - \mathbf{s}_{fid}), \quad (4.2)$$

where  $\widehat{\mathbf{C}}^{-1}$  is an unbiased estimator of the precision matrix. We compute (and report) results based both on the precision matrix estimated at the fiducial model, and the precision matrix estimated at each specific configuration. We used the prescription from [151] to de-bias the estimator of the precision matrix. For each observable, we report the  $\chi^2$  per degree of freedom. Furthermore, to account for the effect of a change in configuration on the covariance matrix, we report two values of the reduced  $\chi^2$ , one based on the covariance matrix for the reference configuration and another based on the configuration-specific covariance. We consider two configurations as statistically equivalent when their  $\chi^2$  per degree of freedom, in the presence of noise, is less than or equal to unity.



### 4.2.3 Hyper-parameter configurations

We analyzed different configurations for two hyper-parameters: the thickness of the lensing planes used in the ray-tracing algorithm, and the mass resolution (number of particles) of the N-body simulations. For each configuration, we ray-traced 10,048 convergence maps, each of these maps with an area of  $12.25 \text{ deg}^2$ .

The positions of the DM particles were saved at redshifts that allowed the construction of light-cones with lens planes at comoving distances of  $20 h^{-1}$ ,  $40 h^{-1}$ ,  $60 h^{-1}$ ,  $80 h^{-1}$ , and  $120 h^{-1}$  Mpc. Thinner lens planes can potentially capture more accurately the evolution of the matter density field with redshift. However the number of planes needed to cover a redshift range increases as the plane thickness decreases, and so do the computational and storage requirements for the simulations. In particular, the two tasks that account for the largest increase in computation time are the calculation of the gravitational potential at the planes (solving a 2D Poisson equation in Fourier space) and the computation of the Jacobian matrix that determines the light ray's deflections at each point on the planes [143]. The fiducial case corresponds to a lens plane thickness of  $80 h^{-1}$  Mpc, a value that has been typically used in prior work [159, 42, 121], provides 9 independent lens planes per simulation snapshot (increasing the number of pseudo-independent  $\kappa$  maps that can be generated from a single N-body simulation), and is not large enough to show discreteness effects with lensed galaxies at  $z = 2$  [183].

The minimum angular scale at which cosmological information can be extracted is limited by the depth of the survey, which determines the number density of galaxies whose shape can be measured, and the accuracy of the forward models used to predict the signal. Baryonic physics [197] and intrinsic alignments [198] restrict the accuracy of current models at small scales. Matching the mass resolution of the underlying N-body simulations to the scales at which the analysis of the data is reliable saves computational resources.

In our fiducial N-body simulations we used  $1024^3$  particles, and we ran additional simulations with  $128^3$ ,  $256^3$  and  $512^3$  particles, which yield mass resolutions per DM particle of  $5.7 \times 10^{12} h^{-1} \text{M}_\odot$ ,  $7.2 \times 10^{11} h^{-1} \text{M}_\odot$ ,  $9.0 \times 10^{10} h^{-1} \text{M}_\odot$  and  $1.1 \times 10^{10} h^{-1} \text{M}_\odot$ . We note that due

	Task	RAM	CPU Time	Change CPU [%]
Plane thickness				
20 $h^{-1}$ Mpc	lens planes	96 GB	82.7	+1450%
40 $h^{-1}$ Mpc			21.3	+300%
60 $h^{-1}$ Mpc			9.3	+75%
<b>80<math>h^{-1}</math> Mpc</b>			<b>5.3</b>	
120 $h^{-1}$ Mpc			2.7	-50%
20 $h^{-1}$ Mpc	Ray tracing	384 GB	1621.3	+407%
40 $h^{-1}$ Mpc			640	+100%
60 $h^{-1}$ Mpc			426.7	+33%
<b>80<math>h^{-1}</math> Mpc</b>			<b>320</b>	
120 $h^{-1}$ Mpc			213.3	-33%
# of particles				
128 <sup>3</sup>	N-body	576 GB	736	-99%
256 <sup>3</sup>		576 GB	1408	-98%
512 <sup>3</sup>		576 GB	6144	-94%
<b>1024<sup>3</sup></b>		<b>9216 GB</b>	<b>103833</b>	

Table 4.1: Memory and computational time requirements for main simulation tasks for different hyper-parameter values. Each CPU time unit is a core hour (representing wall clock time if computed in series). Changes in CPU time are relative to the fiducial run (in bold).

to memory limitations, we increased the number of CPUs allocated to our higher-resolution simulation, resulting in a worse scaling than the one that can be achieved in an optimized setup [193].

Tables 4.1 and 4.2 summarize the computational cost of the main steps involved in our simulation pipeline, and the disk space required for storing the different data products (in practice, not all need to be saved). Performance benchmarks are based on runs using Intel Knights Landing nodes from TACC’s Stampede2 supercomputer at the NSF XSEDE facility.

#### 4.2.4 Observables

We analyze the effect of changes in the hyper-parameters used to generate the synthetic data on the convergence power spectrum and a series of non-Gaussian observables: the one-point probability density function (PDF), lensing peak counts, and the full set of Minkowski functionals (MFs).

The power spectrum is the Fourier transform of the two-point correlation function, encodes all

	Snapshot size	# of snapshots	Total memory	Plane size	# of planes	Total memory
Plane thickness						
20 $h^{-1}$ Mpc		226	792 GB		8064	524.2 GB
40 $h^{-1}$ Mpc		114	400 GB		2016	131 GB
60 $h^{-1}$ Mpc	28.8 GB	77	270 GB	65 M	900	58.5 GB
<b>80 <math>h^{-1}</math> Mpc</b>		<b>58</b>	<b>204 GB</b>		<b>504</b>	<b>32.8 GB</b>
120 $h^{-1}$ Mpc		40	141 GB		228	14.8 GB
# of particles						
128 <sup>3</sup>	55 M		3.2 GB			
256 <sup>3</sup>	448 M		26 GB			
512 <sup>3</sup>	3.5 GB	58	204 GB	65 M	504	32.8 GB
<b>1024<sup>3</sup></b>	<b>28.8 GB</b>		<b>1670.4 GB</b>			

Table 4.2: Individual plane and snapshot size and the respective total storage requirements for both intermediate data products. In bold, values for the fiducial configuration.

the information available in a Gaussian random field, and can be accurately predicted from theory on large angular scales. Therefore, it is commonly used in WL analyses [199, 200]. We measured it in 20 logarithmic bins between  $\ell_{min} = 200$  and  $\ell_{max} = 12,000$ , extending slightly above the smoothing scale of 1 arcmin which corresponds to  $\ell \approx 10,800$ . The relatively featureless power spectra can be properly characterized with 20 bins, without yielding a very large data vector when combining the 5 auto-spectra with the 15 cross-spectra for all 5 redshift bins. Since the synthetic maps cover a small field of view, we can use the flat-sky approximation to model the expected power spectrum from measuring the ellipticities of lensed sources at a fixed radial comoving distance,  $\chi_s$ :

$$P_\kappa(\ell) = \left( \frac{3H_0^2}{2c^2} \Omega_m \right)^2 \int_0^{\chi_s} \frac{d\chi}{a^2(\chi)} \left( 1 - \frac{\chi}{\chi_s} \right)^2 P_\delta \left( k = \frac{\ell}{\chi}; \chi \right). \quad (4.3)$$

Here  $P_\delta$  is the (non-linear) matter power spectrum, which depends on the wavenumber  $k$  and evolves with redshift, and  $H_0$ ,  $\Omega_m$  and  $c$  are the Hubble constant, matter density parameter and speed of light, respectively. Shape noise contributes white noise with total power of  $P_\epsilon = \sigma_\epsilon^2/n_g$ . Smoothing with a Gaussian kernel of width  $\sigma_s$  multiplies the convergence power spectrum by the beam function  $b_\ell = \exp[-\ell(\ell+1)\sigma_s^2]$ .

The effect that the finite thickness of the lens planes has on the measured convergence power spectrum can be modelled as convolving the matter density power spectrum with a window function. Following [201], it suffices to substitute  $P_\delta$  in Eq. (4.3) by

$$P'_\delta(k; \chi) = \frac{\Delta\chi}{\pi} \int_0^{k_\parallel^{\max}} dk_\parallel P_\delta(k) \text{sinc}^2\left(\frac{k_\parallel \Delta\chi}{2}\right), \quad (4.4)$$

where  $k = \sqrt{k_\parallel^2 + k_\perp^2}$ ,  $k_\perp = \ell/\chi$ ,  $\text{sinc}(x) = \sin x/x$ , and  $\Delta\chi$  is the thickness of the lens planes used to build past light-cones in our simulations.

Changes in the mass resolution of the N-body simulations will also affect the measured  $\kappa$  power spectrum. On one hand, the matter power will be suppressed on scales below the simulation resolution. On the other hand, an additional shot-noise component will result from the discreteness of the simulated matter density field. Both effects can be accounted for as in [184], by smoothing the matter power spectrum with a Gaussian kernel of width  $\sigma_N$  and adding a Poisson noise contribution corresponding to the finite number density  $n_p$  of simulated particles,

$$P''_\delta(k; \chi) = P_\delta(k; \chi) \exp\left(-\sigma_N^2 k^2\right) + \frac{1}{n_p}. \quad (4.5)$$

Most scales in our simulated maps probe non-linear structures, and as a result, there is non-Gaussian information not captured by the convergence power spectrum. To assess the impact of the different simulation parameters on the non-Gaussian information content of convergence maps, we analyze the effect on the one-point probability density function, the distribution of lensing peaks, and the Minkowski functionals.

The convergence PDF has been shown to help break the power spectrum degeneracy in the  $\Omega_m - \sigma_8$  plane, improving cosmological constraints from the power spectrum alone (e.g. [202, 203]). It has also been used in order to assess the ability of weak lensing measurements to infer the neutrino masses [204]. We use the binned PDF measured on our (smoothed) simulated maps within the range  $[-3\sigma_\kappa, 5\sigma_\kappa]$ , where  $\sigma_\kappa$  is the standard deviation of the maps corresponding to the highest mass resolution ( $1024^3$  particles), and the fiducial,  $80 h^{-1}$  Mpc, lens planes. For each

redshift bin in the noiseless case, we use  $\sigma_\kappa = [0.007, 0.014, 0.019, 0.023, 0.027]$ , and for the maps that include shape noise  $\sigma_\kappa = [0.034, 0.031, 0.035, 0.043, 0.055]$ . We compare the measurements over simulations with the analytic expectation for Gaussian random fields (GRFs) with variance

$$\sigma_0^2 = \int_0^\infty \frac{\ell d\ell}{2\pi} P_\kappa(\ell). \quad (4.6)$$

Lensing peaks are local maxima of the convergence field; their distribution as a function of their height,  $\kappa$ , was proposed to constrain cosmological parameters [32, 31], and has been used extensively and successfully ever since [47, 48, 53, 52]. We measured peak histograms in 50 linear bins covering the range  $\kappa \in [-0.5, 5.0]$  in units of the r.m.s. shape noise in each redshift bin. For the maps smoothed on 1 arcmin scales, the r.m.s. of the shape noise for the redshift bins [0.5, 1.0, 1.5, 2.0, 2.5] is  $\sigma_n = [0.033, 0.027, 0.030, 0.036, 0.048]$ . As we did for the PDF, we compare measurements with the predictions for GRFs with the same convergence power spectrum as the simulated  $\kappa$  maps. The analytic expressions for the number of peaks in GRFs was derived in [205]. We reproduce here the relevant equations for convenience. The number of lensing peaks in a GRF for a given height  $\nu = \kappa/\sigma_0$  is given by:

$$n_{\text{peaks}}(\nu) = \frac{1}{(2\pi)^{3/2} \theta^2} \exp\left(-\frac{\nu^2}{2}\right) G(\gamma, \gamma\nu), \quad (4.7)$$

where  $\theta = \sqrt{2}\sigma_1/\sigma_2$  and  $\gamma = \sigma_1^2/(\sigma_0\sigma_2)$  depend on the moments of the convergence power spectrum

$$\sigma_p^2 \int_0^\infty \frac{\ell d\ell}{2\pi} \ell^{2p} P_\kappa(\ell) \quad (4.8)$$

and

$$G(\gamma, x) = (x^2 - \gamma^2) \left[ 1 - \frac{1}{2} \operatorname{erfc} \left( \frac{x}{\sqrt{2(1-\gamma^2)}} \right) \right] + x(1-\gamma^2) \frac{\exp\left(-\frac{x^2}{2(1-\gamma^2)}\right)}{\sqrt{2\pi(1-\gamma^2)}} + \frac{\exp\left(-\frac{x^2}{3-2\gamma^2}\right)}{\sqrt{3-2\gamma^2}} \left[ 1 - \frac{1}{2} \operatorname{erfc} \left( \frac{x}{\sqrt{2(1-\gamma^2)(3-2\gamma^2)}} \right) \right]. \quad (4.9)$$

Finally, the three Minkowski functionals that can be defined in a two-dimensional random field,  $\{V_0, V_1, V_2\}$ , measure the area, length, and genus of the subset of points in the random field whose value exceeds a given threshold [150, 206]. While  $V_0$  conveys the same information as the PDF, we include it in our measurements for completeness. They have been extensively used to measure or extract non-Gaussian information from cosmological datasets, including weak lensing [40, 42, 49, 207, 44].

We measured the three MFs in 28 linear bins covering the range  $\kappa \in [-2, 5]$  in units of the r.m.s. shape noise for each redshift bin, and compared it with analytical predictions for GRFs. The MFs of two-dimensional GRFs are determined by the moments of their power spectrum, and their expression as a function of  $\nu$  can be found in [208]:

$$V_0(\nu) = \frac{1}{2} \operatorname{erfc} \left( \frac{\nu}{\sqrt{2}} \right) \quad (4.10)$$

$$V_1(\nu) = \frac{1}{8\sqrt{2}} \frac{\sigma_1}{\sigma_0} \exp \left( -\frac{\nu^2}{2} \right) \quad (4.11)$$

$$V_2(\nu) = \frac{1}{(2\pi)^{3/2}} \left( \frac{\sigma_1}{\sqrt{2}\sigma_0} \right)^2 \nu \exp \left( -\frac{\nu^2}{2} \right). \quad (4.12)$$

While a perturbative expansion has been found for MFs for non-Gaussian random fields [40], they have been shown to converge too slowly at the relevant  $\sim \text{arcmin}$  smoothing scales [42], so we

do not investigate them here.

### 4.3 Results and discussion

We discuss the effect on lensing statistics of the comoving distance between lens planes (plane thickness) in § 4.3.1 and mass resolution in § 4.3.2.

#### 4.3.1 Lens plane thickness

We show the mean percentage difference in the power spectrum for noiseless and noisy convergence maps for different lens plane thicknesses (or comoving distance between lens planes) relative to the fiducial case of  $80 h^{-1}$  Mpc in Figure 4.2. For each plane thickness, the power spectrum is measured and averaged over 10,048 convergence maps. For clarity, we do not display the results for the 10 cross-spectra between redshift bins, but the conclusions are not altered. The gray band in each panel represents a standard error corresponding to 3 times the standard deviation measured over the fiducial maps, scaled to a  $2 \times 10^4 \text{ deg}^2$  LSST-like survey. This does not incorporate the effect of off-diagonal terms in the covariance matrix, which are fully utilized (including auto and cross spectra) in the  $\chi^2$  statistic described in § 4.2.2 to assess the significance of the differences between each configuration and the fiducial case.

The effect of the finite width of lens planes, measured on the simulations, matches well the expectation from the model described in [201], in which the matter power spectrum is convolved with the window function corresponding to a spherical lens shell. Since the differences between the fiducial and the  $60 h^{-1}$  Mpc case are in general sub-percent level, they fall in the linear range of the y-scale and appear to be larger than those for other cases, but they are not. As the lens planes cut structures, there is an appreciable loss of power for thin planes throughout the multipole range considered in this study. The presence of shape noise renders those differences statistically insignificant for the range of thicknesses  $60 h^{-1} - 120 h^{-1}$  Mpc (see Table 4.3).

The effect on the PDF is shown in Figure 4.3, where the PDF change measured from maps built using different lens plane thicknesses is plotted (thick lines) together with predictions for Gaussian

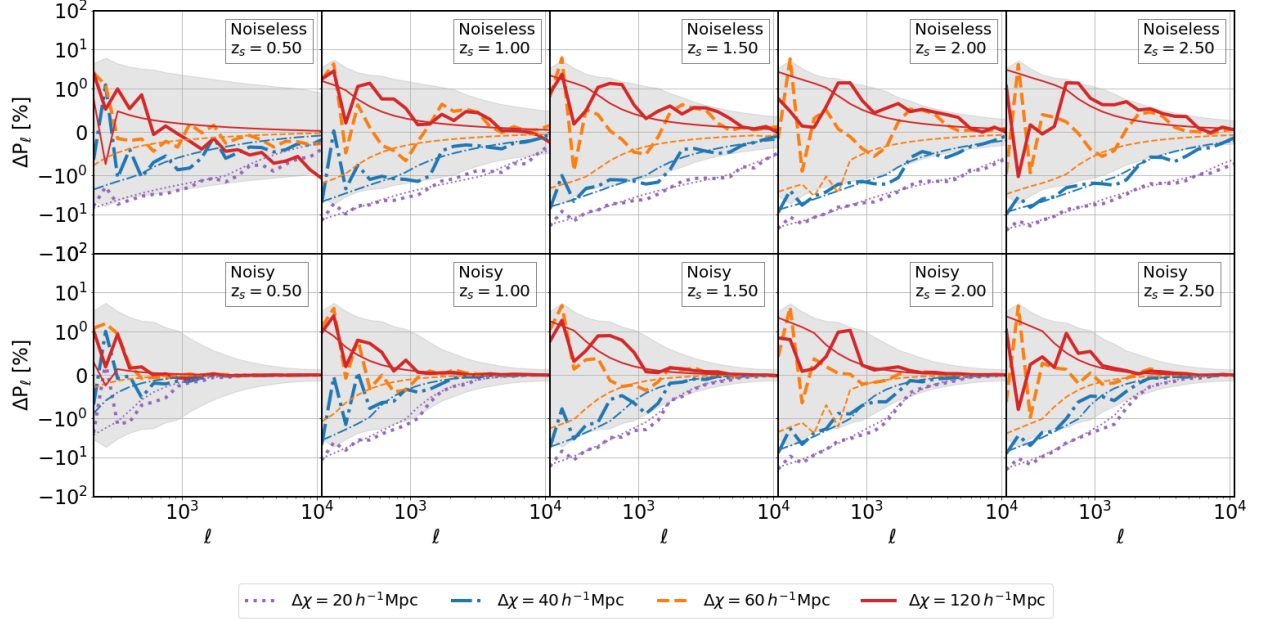


Figure 4.2: Effect of lens plane thickness on convergence power spectra. Each column shows, for a different redshift bin, the percentage difference in the mean auto power spectrum measured over 10,048 convergence maps for a given thickness of the lens planes, relative to the fiducial value of  $80 \ h^{-1} \text{Mpc}$ . The top row corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines represent measurements over simulated maps, while thin lines represent theoretical predictions following [201]. For comparison, a standard error is shown in shaded gray, corresponding to 3 standard deviations of the measurements in the fiducial case, scaled to a  $2 \times 10^4 \text{ deg}^2$  survey.



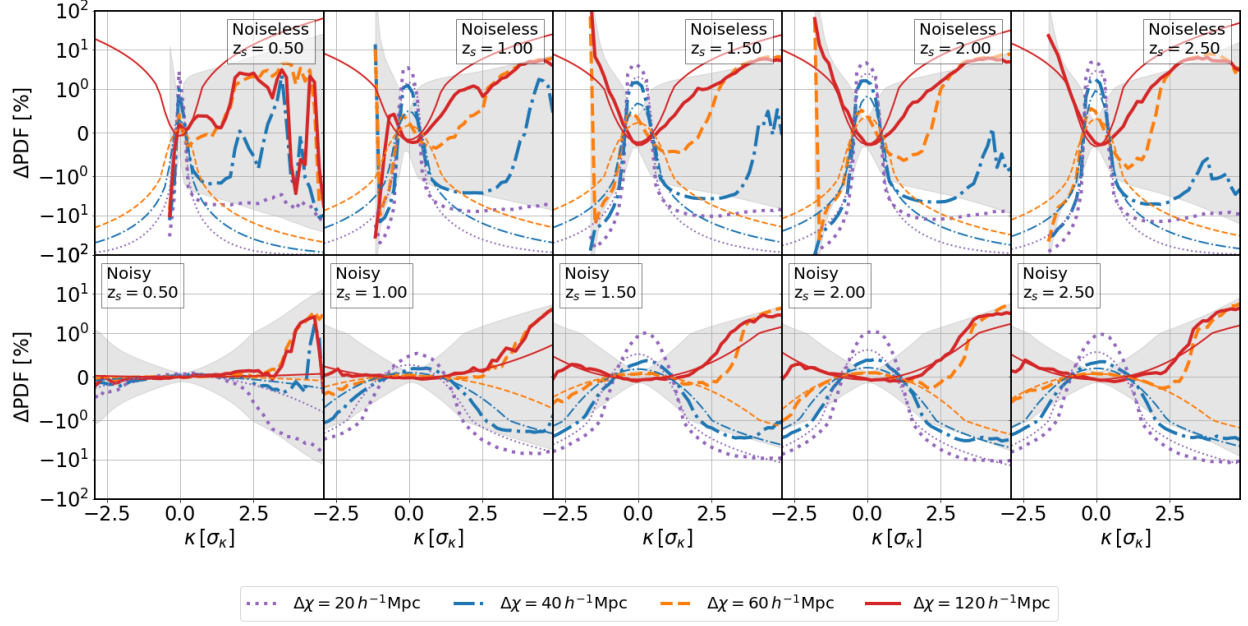


Figure 4.3: Same as Figure 4.2 for the one-point  $\kappa$  probability density function (PDF). Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps.

random fields (thin lines). The GRFs have the same power spectrum as the simulated  $\kappa$  maps. The overall impact on the GRFs is straightforward: the loss in power translates into a smaller variance for the random field (see Eq. (4.2.4)). As a result we would expect, for planes thinner than the fiducial, a suppression in the tails of the PDF compensated by an enhancement near the peak of the PDF (and the opposite for the case with thicker lens planes). Qualitatively, the measurements in the presence of shape noise follow that trend, which is to be expected since the shape noise model considered is Gaussian. In the limit of noiseless data, while there is still some qualitative agreement, the quantitative differences with the GRF model are evident.

From a practical perspective, lens planes whose thickness lays in the range of  $60 - 120 h^{-1}$  Mpc yield convergence maps with statistically indistinguishable PDFs, even in the absence of shape noise (see Table 4.3.1).

The sensitivity of the distribution of lensing peak counts to the width of the lens planes used during raytracing is shown in Figure 4.4. For noiseless data, lens planes thinner than  $60 h^{-1}$  Mpc

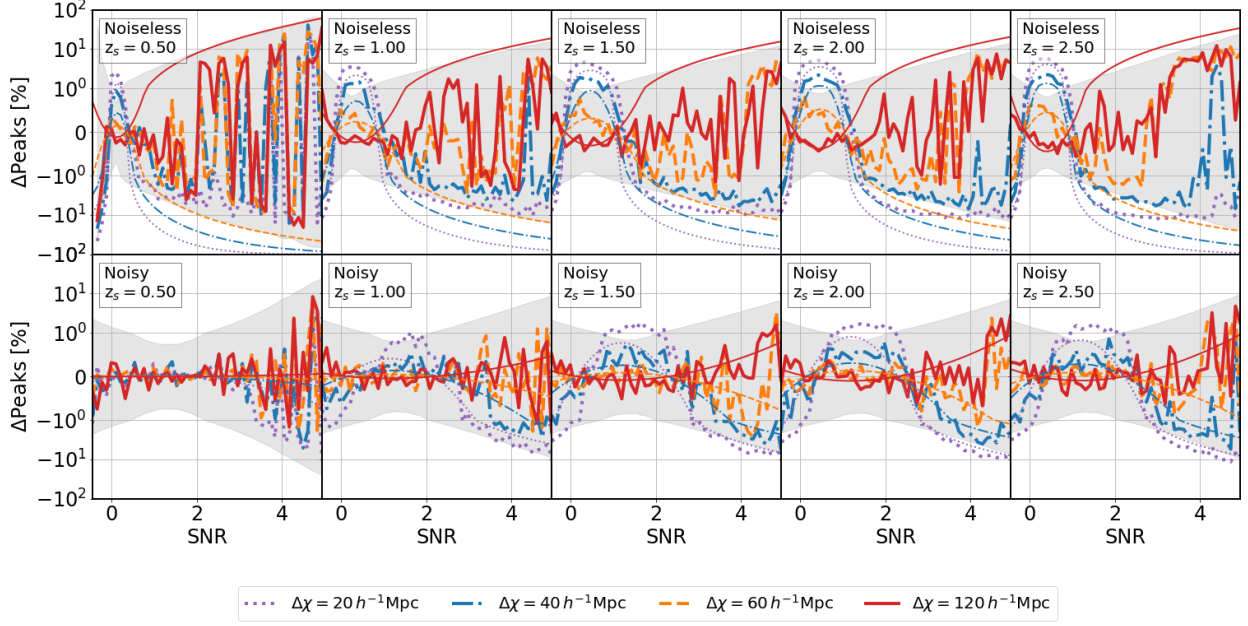


Figure 4.4: Same as Figure 4.2 for the lensing peak distribution. Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps.

overproduce low-significance peaks while underproducing higher-significance ones. Table 4.3 shows how the results from using lens planes within the range  $60 - 120 h^{-1} \text{ Mpc}$  are statistically indistinguishable. The addition of shape noise reduces the differences further, to the point where using thinner planes, at  $40 h^{-1} \text{ Mpc}$ , could be safe. As seen from the power spectrum, there is not a strong case to move from the fiducial thickness of  $80 h^{-1} \text{ Mpc}$ .

As for the PDF, the changes in the lensing peak counts in  $\kappa$  maps are qualitatively similar to those in GRFs with the same power spectrum, but there are quantitative differences. The effect is also similar to the one seen for the PDF, and can be understood intuitively: thicker lens planes provide more power on larger scales that modulates the heights of local maxima, increasing the variance of their distribution (and conversely for thinner planes).

Past studies have found that the use of the Born approximation, in which the convergence is directly estimated weighting the projected matter density, can induce a significant bias to predictions of non-Gaussian observables such as the skewness and kurtosis [142]. Since in the limit of very

thick planes, the multi-plane ray-tracing algorithm is similar to the Born approximation (except for the lack of redshift evolution in the matter density field within each plane), any attempt to predict peak histograms beyond the range of thickness explored in this study, would need to be validated.

The effect of lens plane thickness on the MFs is different for each functional (see Figure 4.5). As expected, shape noise "gaussianizes" the convergence field and as a result measured differences follow qualitatively the predictions for GRFs. This is not the case in the limit of noiseless data. The  $\kappa$  field is less sensitive to changes in the plane thickness than what would be expected for GRFs, especially  $V_1$ . For  $V_2$ , the region of  $\kappa \approx 0$ , where the functional changes sign, yields noisy measurements.

In terms of  $\chi^2$  (Table 4.3), the MFs are more demanding than other non-Gaussian observables. While in the presence of shape noise, lens planes in the range of  $60 - 120 h^{-1}$  Mpc yield measurements that are statistically equivalent, that is not the case anymore in the limit of noiseless data for  $V_1$  and  $V_2$ .

As explained in § 4.2.2, the reduced  $\chi^2$  allows us to assess the impact on inference, of the differences in the measured statistics from synthetic maps generated with different hyper-parameters. Changes in the covariance matrix are taken into consideration when computing  $\chi^2$ . We use either the covariance matrix from maps in the reference configuration, or an alternative covariance from maps recomputed after changing the thickness of the lens planes. In both cases, we use the full covariance to compute  $\chi^2$ , including, for instance, the 10 cross power spectra not shown in Figure 4.2. As an illustration, we display the full covariance matrix for the power spectrum measured on noisy maps in Figure 4.6.

The effect of lens planes in the covariance is small (for example, less than 20% for the dominant, diagonal elements for the power spectrum). As a result, the conclusions are the same regardless of which covariance is used. Also, [191] have analyzed the impact of non-Gaussian contributions to the lensing power spectrum covariance on parameter inference. They find that mean changes in the covariance matrix of  $\approx 20\%$  translate to  $\leq 5\%$  changes in the parameter uncertainties.

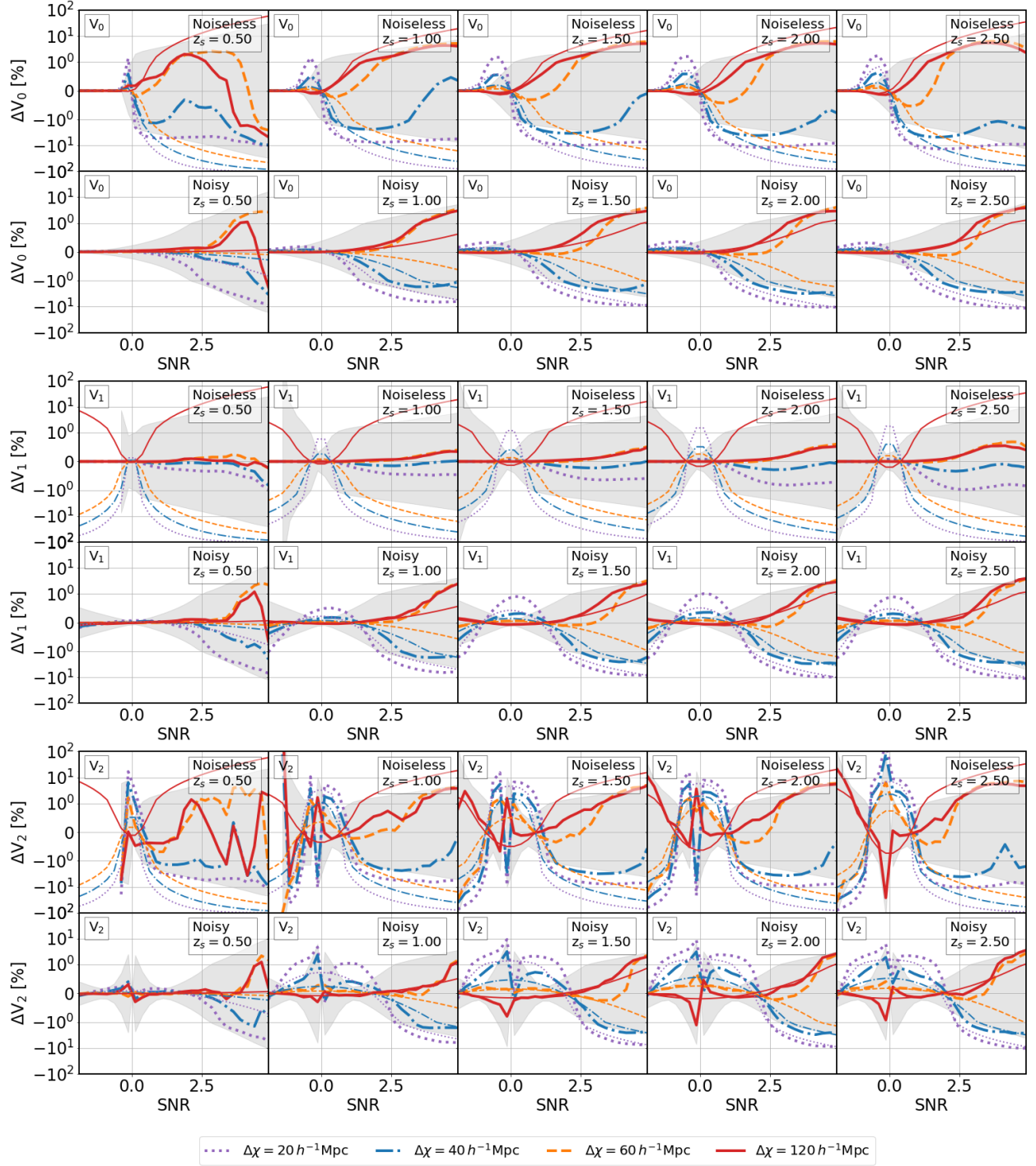


Figure 4.5: Same as Figure 4.2 for the three Minkowski functionals (MFs). Each pair of rows shows a different MF. The top row of each pair corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps.

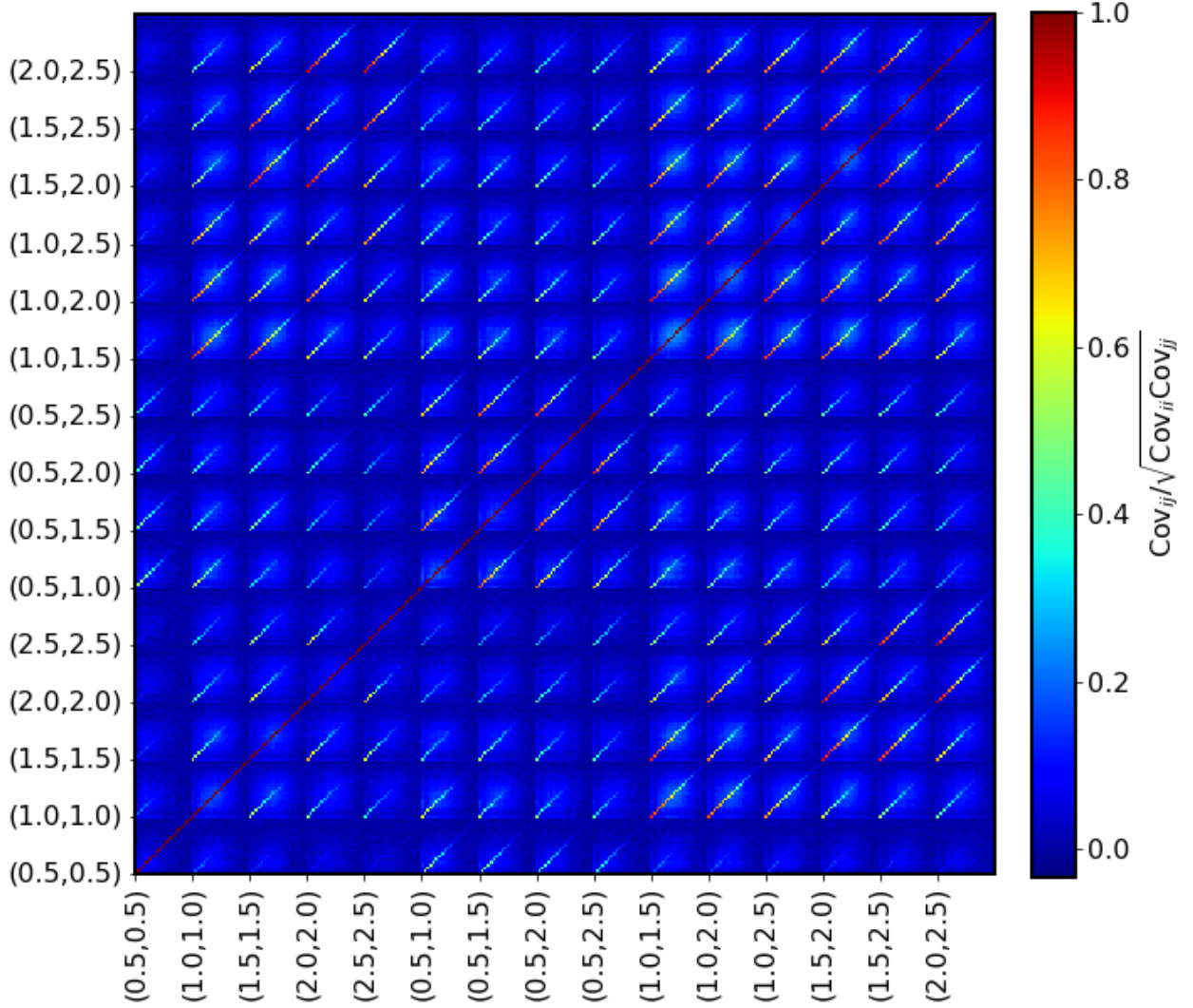


Figure 4.6: Covariance matrix measured over 10048 convergence maps generated from our  $512^3$  particle simulation with  $80 h^{-1}$  Mpc lens planes. The axis labels indicate the redshift bins used. For example, the slice  $[(1.0, 1.5), (2.5, 2.5)]$  shows the correlation between the cross power spectrum of redshift bins  $z_s = 1.0$ ,  $z_s = 1.5$  and the auto power spectrum of the redshift bin  $z_s = 2.5$ .

Table 4.3 collects the results for the observables under study. With the shape noise levels considered, lens planes whose thickness is in the range of  $60 - 120 h^{-1}$  Mpc yield statistically indistinguishable measurements for all the statistics we have studied, and are therefore safe to use. For a scheme, like ours, that recycles N-body snapshots to build different past light-cones, there is a trade-off between a loss in power induced by the lens plane’s window function, and the number of lens planes that can be built from a single N-body simulation. The fiducial choice of  $80 h^{-1}$  Mpc seems reasonable, since it allows to generate as many as a few  $\times 10^4$  pseudo-independent convergence maps from a single simulation [145]. In the limit of noiseless data this is not the case anymore for the power spectrum and MFs. For the power spectrum, the effect of the lens plane thickness can be incorporated analytically.

These results do not take into consideration the uncertainty in the measured covariance matrices: if they are overestimated, our results could change. Nevertheless, the required errors must be substantial: the true covariance would need to be overestimated by more than 150% (its true elements be smaller than  $0.4\times$  the elements of the matrix measured on our simulations) for the Minkowski functionals (the most sensitive statistic to this hyper-parameter) measured on maps generated with  $80 h^{-1}$  Mpc lensing planes to be statistically distinguishable from those measured on maps generated with  $120 h^{-1}$  Mpc lensing planes.

#### 4.3.2 Mass resolution

Figure 4.7 displays the mean percentage difference in the power spectrum for noiseless and noisy maps for different mass resolutions, compared to the highest-resolution case. The number of particles in the four configurations we tested are  $128^3$ ,  $256^3$ ,  $512^3$  and  $1024^3$ . As done in Figure 4.2 with the sensitivity to the lens plane thickness, a standard error corresponding to 3 standard deviations for the fiducial case is shaded for reference in Figure 4.7.

The main difference between the power spectra is an increase in power on small scales, with a relative loss of power on intermediate scales, as the mass resolution decreases. Qualitatively, this behavior matches what is expected from the matter power spectrum: a loss of power at inter-



Thickness	Noiseless		Noisy	
	Model-dependent	Fiducial	Model-dependent	Fiducial
Power spectrum reduced $\chi^2$				
20 $h^{-1}$ Mpc	152.98 (188.31)	98.78 (120.01)	8.23 (11.72)	5.59 (7.95)
40 $h^{-1}$ Mpc	17.33 (16.72)	15.73 (14.74)	<b>0.83</b> (1.17)	<b>0.74</b> (1.04)
60 $h^{-1}$ Mpc	6.96 (4.43)	6.91 (4.34)	<b>0.17 (0.23)</b>	<b>0.17 (0.24)</b>
120 $h^{-1}$ Mpc	9.64 (5.25)	8.84 (5.09)	<b>0.16 (0.21)</b>	<b>0.16 (0.22)</b>
PDF reduced $\chi^2$				
20 $h^{-1}$ Mpc	30.67 (40.27)	18.02 (23.69)	10.25 (13.31)	7.50 (9.78)
40 $h^{-1}$ Mpc	3.09 (3.96)	2.51 (3.20)	1.33 (1.66)	1.23 (1.53)
60 $h^{-1}$ Mpc	<b>0.65 (0.72)</b>	<b>0.62 (0.69)</b>	<b>0.41 (0.46)</b>	<b>0.41 (0.44)</b>
120 $h^{-1}$ Mpc	<b>0.48 (0.52)</b>	<b>0.51 (0.54)</b>	<b>0.38 (0.39)</b>	<b>0.39 (0.39)</b>
Peak histogram reduced $\chi^2$				
20 $h^{-1}$ Mpc	14.50 (17.12)	11.33 (13.52)	5.74 (6.34)	4.78 (5.40)
40 $h^{-1}$ Mpc	2.08 (2.41)	1.87 (2.17)	<b>0.97</b> (1.02)	<b>0.89 (0.95)</b>
60 $h^{-1}$ Mpc	<b>0.61 (0.62)</b>	<b>0.61 (0.60)</b>	<b>0.37 (0.33)</b>	<b>0.36 (0.32)</b>
120 $h^{-1}$ Mpc	<b>0.56 (0.53)</b>	<b>0.56 (0.52)</b>	<b>0.36 (0.37)</b>	<b>0.36 (0.37)</b>
$V_0$ reduced $\chi^2$				
20 $h^{-1}$ Mpc	58.55 (82.06)	32.13 (44.96)	18.16 (25.01)	13.17 (18.27)
40 $h^{-1}$ Mpc	5.38 (7.39)	4.37 (6.01)	2.16 (2.79)	1.95 (2.54)
60 $h^{-1}$ Mpc	<b>0.99</b> (1.19)	<b>0.97</b> (1.16)	<b>0.54 (0.58)</b>	<b>0.54 (0.57)</b>
120 $h^{-1}$ Mpc	<b>0.70 (0.81)</b>	<b>0.81 (0.94)</b>	<b>0.39 (0.41)</b>	<b>0.40 (0.41)</b>
$V_1$ reduced $\chi^2$				
20 $h^{-1}$ Mpc	56.34 (78.72)	41.79 (58.42)	19.70 (26.83)	13.89 (19.10)
40 $h^{-1}$ Mpc	6.61 (9.04)	5.78 (7.93)	2.25 (2.93)	2.05 (2.68)
60 $h^{-1}$ Mpc	1.20 (1.47)	1.16 (1.42)	<b>0.60 (0.56)</b>	<b>0.62 (0.57)</b>
120 $h^{-1}$ Mpc	<b>1.87 (2.40)</b>	<b>2.78 (3.66)</b>	<b>0.41 (0.38)</b>	<b>0.41 (0.37)</b>
$V_2$ reduced $\chi^2$				
20 $h^{-1}$ Mpc	46.94 (64.68)	33.04 (45.63)	18.26 (24.98)	13.32 (18.30)
40 $h^{-1}$ Mpc	5.26 (7.00)	4.53 (6.05)	2.13 (2.79)	1.96 (2.57)
60 $h^{-1}$ Mpc	1.11 (1.25)	1.11 (1.23)	0.61 (0.45)	<b>0.62 (0.44)</b>
120 $h^{-1}$ Mpc	1.43 (1.61)	1.67 (1.94)	<b>0.40 (0.32)</b>	<b>0.40 (0.31)</b>

Table 4.3: Goodness of fit for different lens plane thickness configurations, based on the reduced  $\chi^2$  (i.e.  $\chi^2$  per degree of freedom). Configurations that yield good fits ( $\chi^2 \leq 1$ ), implying that they are indistinguishable from the fiducial case, are highlighted in bold. **Power spectrum:** values for a range of  $\ell \in [200, 12000]$  and in parenthesis  $\ell \in [200, 3532]$ . **PDF:** values for a range of  $\kappa \in [-3.0, 5.0]$  in units of the shape noise r.m.s., and in parenthesis  $\kappa \in [-3.0, 3.1]$ . **Peak counts:** values for a range of  $\kappa \in [-0.5, 5.0]$  in units of the shape noise r.m.s., and in parenthesis  $\kappa \in [-0.5, 4.0]$ . **MFs:** values for a range of  $\kappa \in [-2.0, 5.0]$  in units of the shape noise r.m.s., and in parenthesis  $\kappa \in [-2.0, 3.0]$ .

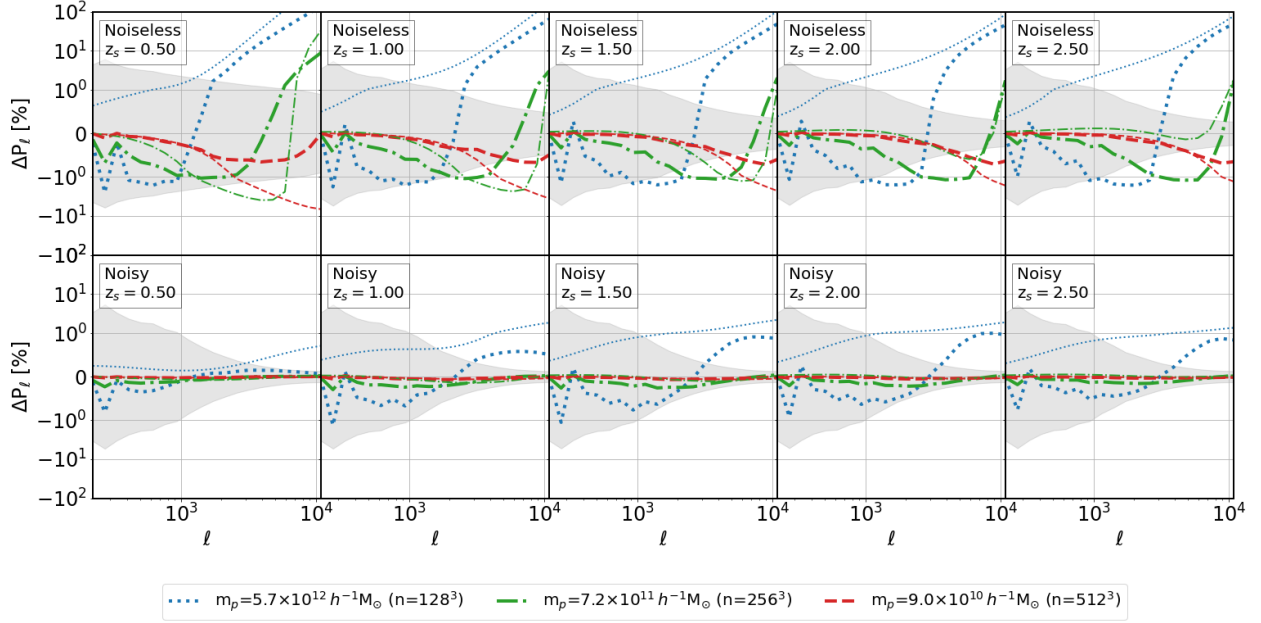


Figure 4.7: Effect of mass resolution (number of particles in the simulation volume) on the convergence power spectrum. Each column shows, for a different redshift bin, the percentage difference in the mean auto power spectrum measured over 10,048 convergence maps for a given mass resolution of the underlying N-body simulation, relative to the fiducial value of  $1.1 \times 10^{10} M_{\odot}$  (corresponding to  $1024^3$  DM particles in the simulation box). The top row corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines represent measurements over simulated maps, while thin lines represent theoretical predictions following Eq. (4.5). For comparison, a standard error is shown in shaded gray, corresponding to 3 standard deviations of the measurements in the fiducial case, scaled to a  $2 \times 10^4 \text{ deg}^2$  survey.



mediate scales as the mass resolution decreases (captured by the first term in Eq. (4.5)) followed by a sharp increase at small scales due to shot noise. Similar effects have been described in past studies [183, 184]. We cannot fit all the data with the simple model in Eq. (4.5), though. For example, a smoothing scale of  $\sigma_N = 5\%$  of the average inter-particle separation gives results that are qualitatively in line with the relative power change when going from  $256^3$  to  $1024^3$  particles, but not for the other resolutions tested. An effect not taken into consideration in that simple model is that of the softening length on the relaxation time of the simulations. All of our simulations kept that length constant, and the softening for the lower resolution runs may have been insufficient. Another effect that can induce a loss in power is the reduction in the linear growth factor due to the discreteness of the initial conditions, as described in [186]. The presence of shape noise, which dominates at small scales, mitigates both effects, rendering the measured statistic more forgiving to the resolution of the underlying N-body simulation.

The effect of mass resolution on the one-point probability density function is not as straightforward as for the lens plane thickness. The GRF model does not work particularly well for noiseless data, as can be seen in Figure 4.8. Except for the lowest-resolution simulation, shape noise still keeps any differences within the standard error of the simulations themselves. In the absence of noise,  $512^3$  and  $1024^3$  particles seem to yield equivalent results.

Lensing peaks are a more robust statistic to the mass resolution (see Figure 4.9), and the configurations with  $256^3$  and  $512^3$  particles are statistically indistinguishable from the  $1024^3$  fiducial for both noisy and noiseless data. The model with the lowest mass resolution yields histograms whose differences from the fiducial case clearly exceed the statistical uncertainty. The difference is more important for low-significance peaks, which can be induced by the Poisson shot noise present in the underlying N-body simulation. Shape noise reduces the differences at the low significance tail, where peaks are noise-dominated. The GRF model for peak counts seems to work better than for the PDF.

The effect of mass resolution on the measured MFs is displayed in Figure 4.10. As with the PDF, a GRF model does not fit as well the impact of changes in mass resolution as it did for changes

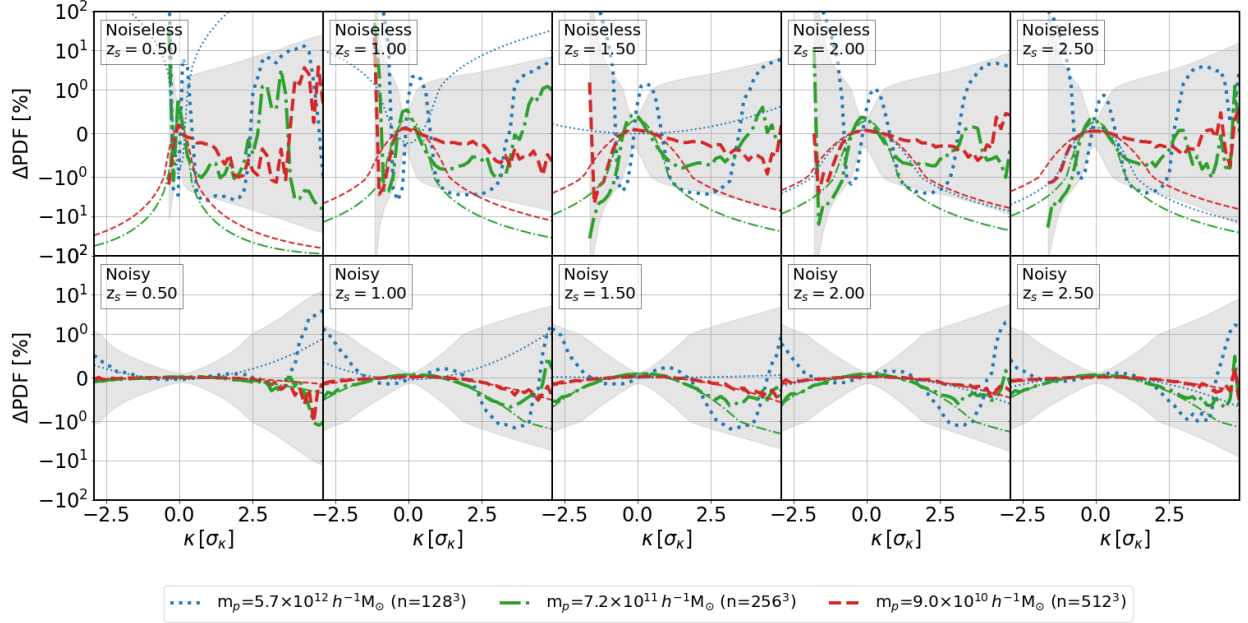


Figure 4.8: Same as Figure 4.7 for the one-point  $\kappa$  probability density function (PDF). Thick lines represent measurements over simulated maps, while thin lines represent the expectation for GRFs with the same power spectrum as the  $\kappa$  maps.

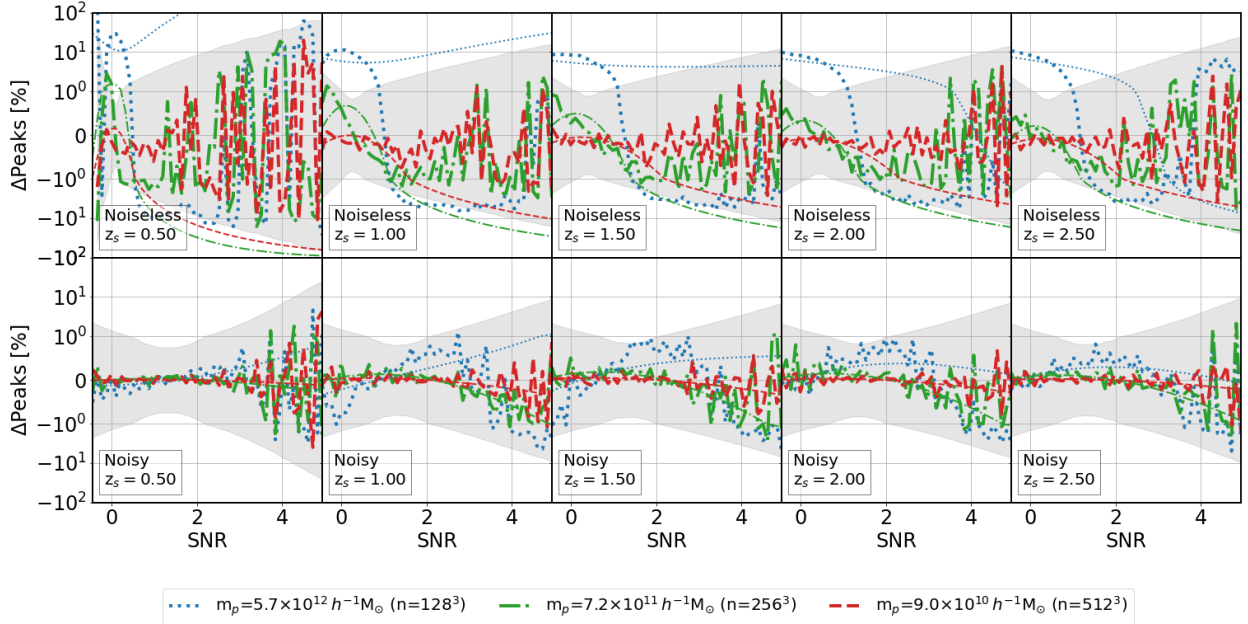


Figure 4.9: Same as Figure 4.2 for the lensing peak distribution. Thick lines represent measurements over simulated maps, while thin lines represent the expectation for GRFs with the same power spectrum as the  $\kappa$  maps.

in lens plane thickness. This may be partly due to the relatively larger change in power, as well as the  $\ell$ -dependence of that change. In general, measurements performed on simulations are less sensitive than what would be expected from GRFs, and their changes remain within errors for most cases.

Table 4.4 collects the statistical significance of changes in the mass resolution of the N-body simulations used to build synthetic  $\kappa$  maps on the observables under study. Most of the considerations discussed in § 4.3.1 about differences in the covariance matrices for the observables apply here as well. For example, the differences for the power spectrum covariance are modest with differences in the dominant, diagonal terms, of less than 3% between the covariances for the  $256^3$  and  $1024^3$  particle simulations. For analyses in the presence of shape noise levels commensurate with the ones considered in this study,  $256^3$  particles suffice. If the multipole range is limited to  $\approx 3500$ , even lower resolutions can be used for the power spectrum (or alternatively the same  $256^3$  resolution can be used for negligible levels of shape noise). The non-Gaussian statistics that are the most sensitive to resolution effects are the MFs.

As with the results for the thickness of the lensing planes, the figures in Table 4.4 do not incorporate the possible effect of errors in the covariance matrices used in the calculation of  $\chi^2$ . In this case, the Minkowski functionals are also the most sensitive statistic. For their measurements on the  $256^3$  particles' simulation to be statistically different than the measurements on the  $1024^3$  particles' maps, the true covariance would need to be overestimated by more than 117%.

## 4.4 Conclusions

We performed a series of numerical experiments to test the influence of the lens plane thickness and the mass resolution of ray-traced N-body simulations on commonly used WL statistics: the convergence power spectrum, the one-point probability density function, lensing peak counts, and Minkowski functionals. While our simulations cannot be used directly to analyze survey data, they can serve to guide design choices in studies of non-Gaussian statistics on small scales. They set some minimal requirements to avoid significant biases in predictions obtained from the application

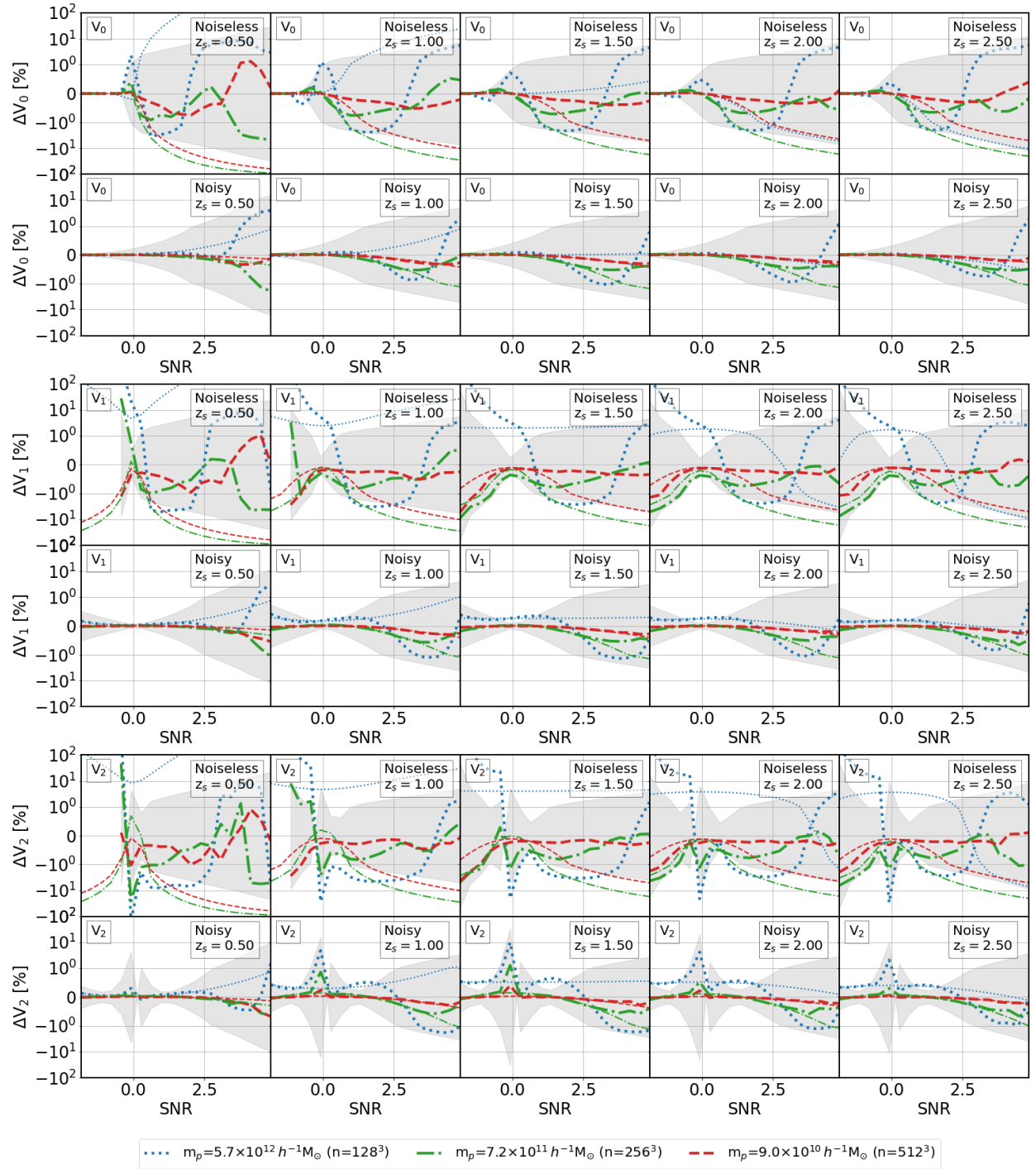


Figure 4.10: Same as Figure 4.2 for the three Minkowski functionals (MFs). Each pair of rows shows a different MF. The top row of each pair corresponds to noiseless data, and the bottom row to data in the presence of shape noise (in all cases, maps were smoothed at 1 arcmin resolution). Thick lines correspond to measurements over simulated data, thin lines to predictions for Gaussian random fields with the same power spectrum as the simulated maps.

Particle mass (number or particles)	Noiseless		Noisy	
	Model-dependent	Fiducial	Model-dependent	Fiducial
Power spectrum reduced $\chi^2$				
$5.7 \times 10^{12} h^{-1} M_{\odot} (128^3)$	11494.43 (25.03)	44822.28 (25.69)	42.44 ( <b>0.85</b> )	44.26 ( <b>0.85</b> )
$7.2 \times 10^{11} h^{-1} M_{\odot} (256^3)$	48.97 ( <b>0.63</b> )	50.48 ( <b>0.62</b> )	<b>0.17 (0.10)</b>	<b>0.17 (0.10)</b>
$9.0 \times 10^{10} h^{-1} M_{\odot} (512^3)$	1.68 ( <b>0.05</b> )	1.66 ( <b>0.04</b> )	<b>0.03 (0.01)</b>	<b>0.03 (0.01)</b>
PDF reduced $\chi^2$				
$5.7 \times 10^{12} h^{-1} M_{\odot} (128^3)$	67.96 (89.38)	289.44 (382.24)	<b>0.72 (0.85)</b>	<b>0.71 (0.83)</b>
$7.2 \times 10^{11} h^{-1} M_{\odot} (256^3)$	1.33 (1.66)	<b>0.16 (0.16)</b>	<b>0.16 (0.16)</b>	<b>0.16 (0.16)</b>
$9.0 \times 10^{10} h^{-1} M_{\odot} (512^3)$	<b>0.10 (0.06)</b>	<b>0.10 (0.06)</b>	<b>0.05 (0.03)</b>	<b>0.05 (0.04)</b>
Peak histogram reduced $\chi^2$				
$5.7 \times 10^{12} h^{-1} M_{\odot} (128^3)$	404.38 (488.55)	426.60 (513.32)	1.44 (1.68)	1.44 (1.68)
$7.2 \times 10^{11} h^{-1} M_{\odot} (256^3)$	<b>0.74 (0.84)</b>	<b>0.76 (0.86)</b>	<b>0.24 (0.24)</b>	<b>0.24 (0.24)</b>
$9.0 \times 10^{10} h^{-1} M_{\odot} (512^3)$	<b>0.29 (0.31)</b>	<b>0.28 (0.30)</b>	<b>0.07 (0.07)</b>	<b>0.08 (0.07)</b>
$V_0$ reduced $\chi^2$				
$5.7 \times 10^{12} h^{-1} M_{\odot} (128^3)$	53.02 (70.59)	89.27 (119.06)	1.47 (1.71)	1.47 (1.70)
$7.2 \times 10^{11} h^{-1} M_{\odot} (256^3)$	<b>0.68 (0.76)</b>	<b>0.69 (0.77)</b>	<b>0.16 (0.19)</b>	<b>0.16 (0.20)</b>
$9.0 \times 10^{10} h^{-1} M_{\odot} (512^3)$	<b>0.07 (0.05)</b>	<b>0.07 (0.05)</b>	<b>0.02 (0.02)</b>	<b>0.02 (0.02)</b>
$V_1$ reduced $\chi^2$				
$5.7 \times 10^{12} h^{-1} M_{\odot} (128^3)$	1273.03 (1691.45)	2103.48 (2796.06)	8.83 (11.26)	8.81 (11.29)
$7.2 \times 10^{11} h^{-1} M_{\odot} (256^3)$	15.19 (20.21)	16.15 (21.47)	<b>0.46 (0.58)</b>	<b>0.46 (0.58)</b>
$9.0 \times 10^{10} h^{-1} M_{\odot} (512^3)$	<b>0.70 (0.89)</b>	<b>0.69 (0.88)</b>	<b>0.07 (0.08)</b>	<b>0.07 (0.08)</b>
$V_2$ reduced $\chi^2$				
$5.7 \times 10^{12} h^{-1} M_{\odot} (128^3)$	5884.23 (7837.70)	13038.65 (17416.22)	4.37 (5.46)	4.40 (5.51)
$7.2 \times 10^{11} h^{-1} M_{\odot} (256^3)$	126.88 (168.00)	137.14 (182.01)	<b>0.29 (0.34)</b>	<b>0.29 (0.35)</b>
$9.0 \times 10^{10} h^{-1} M_{\odot} (512^3)$	<b>0.84 (1.08)</b>	<b>0.83 (1.07)</b>	<b>0.04 (0.04)</b>	<b>0.04 (0.04)</b>

Table 4.4: Goodness of fit as in Table 4.3, but for different mass resolution configurations. Configurations with  $\chi^2 \leq 1$ , implying that they are indistinguishable from the fiducial case, are highlighted in bold. **Power spectrum:** values for a range of  $\ell \in [200, 12000]$  and in parenthesis  $\ell \in [200, 3532]$ . **PDF:** values for a range of  $\kappa \in [-3.0, 5.0]$  in units of the shape noise r.m.s., and in parenthesis  $\kappa \in [-3.0, 3.1]$ . **Peak counts:** values for a range of  $\kappa \in [-0.5, 5.0]$  in units of the shape noise r.m.s., and in parenthesis  $\kappa \in [-0.5, 4.0]$ . **MFs:** values for a range of  $\kappa \in [-2.0, 5.0]$  in units of the shape noise r.m.s., and in parenthesis  $\kappa \in [-2.0, 3.0]$ .

of the multiple lens plane algorithm.

We have found that in the multi-plane ray-tracing algorithm, lens planes in the range  $60 - 120 h^{-1} \text{ Mpc}$  are safe to use in analysis of WL data with galaxy densities commensurate with those in LSST-like surveys. Thinner planes will result in a loss of power across a wide range of multipoles, that can in principle be accounted for analytically for the power spectrum but not for non-Gaussian statistics. On the thick-plane end, there are no biases induced in the observables, and the computational time for raytracing past light cones can be reduced (e.g. by  $\approx 33\%$  when  $120 h^{-1} \text{ Mpc}$  planes are used instead of  $80 h^{-1} \text{ Mpc}$ ). However, such a choice implies a larger number of N-body simulations to generate the same number of pseudo-independent realizations of the  $\kappa$  field, increasing the overall computation budget.

In order to analyze WL data sets at angular resolutions of 1 arcmin with LSST levels of shape noise, simulations with mass resolutions of  $7.2 \times 10^{11} M_{\odot}$  per DM particle are sufficient (corresponding to  $256^3$  particles in a  $240 h^{-1} \text{ Mpc}$  simulation box), even if non-Gaussian statistics are included in the analysis. Moving, for instance from  $512^3$  to  $256^3$  particles could bring computational time savings of  $\approx 77\%$ .

## Chapter 5: Probing gaseous galactic halos through the rotational kSZ effect

### 5.1 Introduction

A deeper understanding of galaxy formation and evolution requires comparing the expected properties of the circumgalactic medium (CGM) with observations. These properties, predicted by simulations, include the CGM’s density, composition, ionization state, and kinematics, as well as the evolution of these quantities through cosmic time. For reviews of the CGM and its connection to galaxy evolution, see, e.g. [209, 210].

Cosmic microwave background (CMB) photons interact with the free electrons in the CGM plasma, and can therefore probe the CGM’s properties. The kinematic Sunyaev-Zeldovich effect (kSZ) is the gain/loss of momentum of these photons as they scatter coherently off electrons with a bulk motion relative to the CMB [211]. The kSZ effect can be used to either learn about the free electron distribution given some kinematic information, or to infer the CGM’s peculiar velocity given its free electron density. We refer the reader to [212] for a detailed review of this effect.

The use of mean pairwise statistics enabled the early detection of the kSZ effect induced by the proper motions of galaxy clusters [213], and the same method has recently been applied successfully to galaxies [214, 215]. The kSZ signal due to clusters’ proper motions has also been detected in stacked data [216, 217] and through high-resolution imaging of individual systems [218]. It has also been detected in cross-correlation analyses of projected fields [219, 220].

Rotating gaseous halos should imprint an additional, dipole-like temperature pattern in the CMB at their location. This signal, which we will refer to as rotational kSZ effect (rkSZ, as in [221]), appears on small angular scales ( $\lesssim 10$  arcmin, corresponding to the halo virial radii  $R_{\text{vir}}$ ), and has been studied in the context of galaxy clusters both analytically [222, 223] and with

simulations [224]. Recently, a tentative detection has been claimed [221], stacking *Planck* data<sup>1</sup> on the location of rotating clusters identified in the Sloan Digital Sky Survey (SDSS,<sup>2</sup>).

In the near future, high-resolution CMB experiments will allow an extension of these studies to probe the rotation of the gaseous halos of individual galaxies. While the signal-to-noise for individual galaxies will remain too low, spin orientations can be estimated for large numbers of nearby galaxies in forthcoming spectroscopic surveys. Motivated by this prospect, in this paper, we assess the feasibility of detecting the rkSZ effect via stacking CMB data on many galaxies. At present, asymmetries in the CMB temperature aligned with the rotation axis of nearby galaxies have been measured [225, 226, 227, 228], but the origin of these asymmetries is not yet fully understood (see § 5.6).

Our manuscript is organized as follows. We start with a description of our models for the rkSZ signal from gaseous galactic halos (§ 5.2). We next describe how to stack CMB data and extract the rkSZ signal statistically (§ 5.3), and forecast the number of galaxies needed for a  $3\sigma$  detection for a variety of experimental settings (§ 5.4). We then proceed to apply these techniques to existing public CMB and galaxy survey data (§ 5.5). In particular, we derive an upper limit on the mean CMB temperature asymmetry in *Planck* data, associated with galaxy spins in the spectroscopic MaNGA survey (Mapping Nearby Galaxies at APO,<sup>3</sup>). Finally, we discuss different caveats and extensions of our analysis and results (§ 5.6) and summarize our main conclusions (§ 5.7).

All calculations assume a flat  $\Lambda$ CDM cosmology with  $\Omega_m = 0.316$ ,  $\Omega_b = 0.048$ ,  $h = 0.675$  and  $T_{\text{cmb}} = 2.725$  K.

## 5.2 Modeling the rotational kSZ (rkSZ) signal from galaxies

### 5.2.1 The rkSZ imprint on the CMB

Free electrons moving relative to the Hubble flow induce temperature anisotropies on the CMB through scattering. This kinetic effect is frequency-independent and cannot be isolated from the

---

<sup>1</sup>*Planck* collaboration:<https://www.cosmos.esa.int/web/planck>

<sup>2</sup>Sloan Digital Sky Survey:<https://www.sdss.org/>

<sup>3</sup>Mapping Nearby Galaxies at APO:<https://www.sdss.org/surveys/manga/>



primordial CMB in the same way as the thermal SZ effect. The kSZ-induced temperature fluctuations depend on the line-of-sight (*los*) integral of the density as well as the peculiar velocity of the electrons. Since the CGM is optically thin to photons from the CMB, we can express the relative change in temperature from free electrons in a galactic halo using the single-scattering limit,

$$\begin{aligned} \frac{\Delta T}{T}(\vec{n}) &= \frac{\sigma_T}{c} \int_{los} dl n_e \vec{v} \cdot \vec{n} \\ &= \frac{\sigma_T}{c} \int_{los} dl n_e(r) v(R) \cos \phi \sin i \end{aligned} \quad (5.1)$$

where  $\vec{n}$  is the unit vector that defines the point on the sky where the CMB temperature is measured,  $\sigma_T$  is the Thomson cross section,  $c$  is the speed of light,  $n_e$  is the electron density, and  $\vec{v}$  is the velocity of the electrons in the CMB rest frame. The last equality applies to a spherically symmetric distribution of free electron ( $r$  is the distance to the halo's center), moving along circular orbits of radius  $R$  with velocity  $v(R)$ . The azimuth angle is  $\phi$  and the galaxy's inclination angle  $i$  (0 deg for a face-on galaxy, 90 deg for edge-on). Fig. 5.3 shows an example of the dipole-like temperature anisotropy induced by a rotating halo on the CMB.

For simplicity, we do not include the kSZ effect induced by the galaxy's mean peculiar velocity in our models, but we discuss its effect on measurements in § 5.6.1 (along with the effect of uncertainties on model parameters, such as the inclination angle and stellar mass of each galaxy).

### 5.2.2 Galactic atmospheres: electron density

The first ingredient needed to estimate the rkSZ signal is the electron density, which, for simplicity, we assume to be spherically symmetric,  $n_e(r)$ .

A simple reference model, which has been used in the study of rotating galaxy clusters, is one with fully ionized hot gas in hydrostatic equilibrium within the gravitational potential of the galaxy's host dark matter (DM) halo [222, 223, 224, 221]. While this model cannot describe galaxy-sized halos, for which a significant fraction of the gas is in a cold and neutral phase, it is still useful as an upper limit to the electron number density. We will refer to such a galactic

atmosphere as “hot”.

A more realistic electron density distribution is given by the multi-phase atmospheric model developed in [229]. We reproduce here its main equations for convenience, and refer the reader to [229] for more detailed explanations. The difference in the distribution of ionized gas in this model and the hot upper limit is shown in Fig. 5.1 for three galaxies of different mass.

The starting point for the multi-phase model is the galaxy’s stellar mass, which we assume to be independent of redshift, that is  $M_\star(z) \approx M_\star(0)$ . This approximation is justified because we only work with galaxies in the local universe. For a given  $M_\star(0)$ , we find the virial mass of the galaxy’s host halo,  $M_v(0)$ , using the fit in [230] (see Eqs. A3-A4 in their appendix). We then scale  $M_v(0)$  to the halo mass  $M_v(z)$  at earlier redshift using the relationship, based on  $N$ -body simulations, in [231]:

$$M_v(z) = M_v(0) \exp\left(-\frac{8.2z}{C_v^0}\right). \quad (5.2)$$

This assumes an NFW profile [173] for the host DM halos. The mass-dependent NFW halo concentration parameter at zero redshift,  $C_v^0$ , is derived from the fit to simulations in [232]:

$$C_v^0 = 9 \left( \frac{M_v(0)}{1.5 \times 10^{13} / h M_\odot} \right)^{-0.13}. \quad (5.3)$$

We define the halo’s virial radius and mass following the equations in [233]. For simplicity, we further assume that the total baryonic mass inside a halo corresponds to its cosmic mass fraction,  $f_b = \Omega_b / \Omega_m$ . We relax this assumption, and discuss how lower baryon fractions affect our results, in § 5.6.1.

The hot atmosphere, used as an upper bound, is fully determined by the DM halo mass and its baryon fraction (see Eqs. 9-11 in [229]). Defining  $\xi \equiv r/r_s$  as the dimensionless radial coordinate normalized by the halo’s scale radius  $r_s \equiv R_{\text{vir}}/C_v^0$ , the free electron density profile for the hot atmosphere,  $n_e^h(\xi)$ , is given by:

$$n_e^h(\xi) = \frac{\rho_0}{\mu_e m_p \left(\xi + \frac{3}{4}\right) (\xi + 1)^2}, \quad (5.4)$$

$$\rho_0 = \frac{f_b M_v}{4\pi r_s^3 g(C_v)}, \quad (5.5)$$

$$g(x) \equiv 9 \ln \left(1 + \frac{4}{3}x\right) - 8 \ln(1+x) - \frac{4x}{1+x}. \quad (5.6)$$

We use a mean atomic weight per electron  $\mu_e = 1.18$  (appropriate for ionized gas with mean cosmological abundance ratios) and  $m_p$  is the proton mass.

In galaxy-sized halos, a significant fraction of the baryons cool and condense into a neutral phase, a part of which form stars. The cooling time of the gas depends on its density, temperature, and cooling rate,  $\Lambda$ . For a halo whose time since its formation is  $t_f$ , the electron density above which hot gas has had time to cool is:

$$n_e^c = \frac{3\mu_e k_b T}{2\mu_i t_f \Lambda(T, Z_g)}, \quad (5.7)$$

where  $k_b$  is the Boltzmann constant,  $T$  is the temperature corresponding to the halo's maximum circular velocity, and  $Z_g$  the metallicity of the gas. The halo's formation time is the lookback time to the redshift at which it has accreted half its mass. We adopt the cooling function parametrized in Appendix A of [229] for a metallicity of  $Z_g = 0.3 Z_\odot$ .

The density in the outer regions of massive halos is below  $n_e^c$ , gas hasn't had time to cool, and the free electron density is given by Eq. 5.4. In the inner regions, the density exceeds  $n_e^c$  and most gas cools into a neutral phase. The transition between the two regimes takes place at the cooling radius. In the inner regions, there will still be some residual hot gas. We will refer to the ionized component of galactic atmospheres as "coronae". Assuming the hot corona reaches hydrostatic equilibrium adiabatically and its density at the cooling radius matches the cooling density, its free

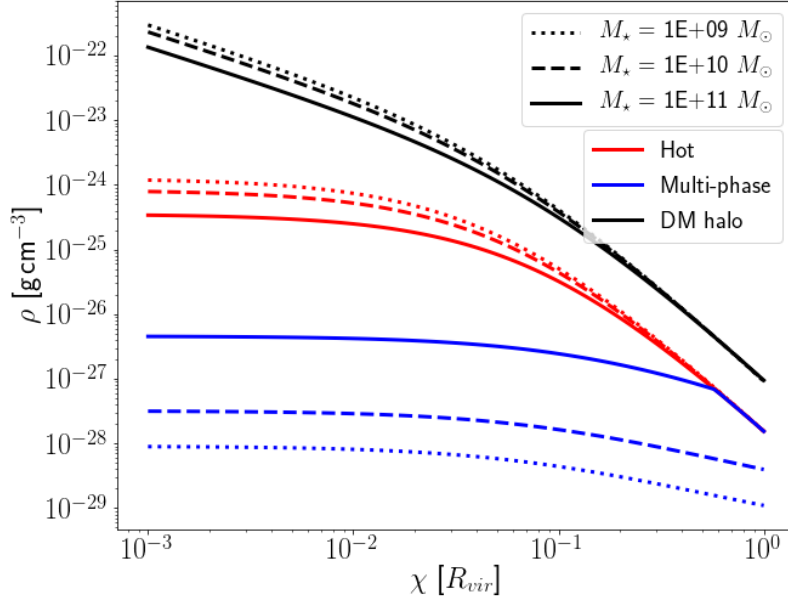


Figure 5.1: Density profiles for the ionized gas for fully ionized galactic atmospheres (“hot”, in red) and multi-phase atmospheres (“multi-phase”, in blue) as a function of distance to the center in virial radii units for galaxies of three different stellar masses. The gas metallicity is  $Z_g = 0.3 Z_\odot$  and the DM density profile of the host halo is shown in black for reference. The halos’ virial masses are  $\{1.2 \times 10^{11}, 3.4 \times 10^{11}, 3.3 \times 10^{12}\} M_\odot$ , their virial radii  $\{69, 98, 215\}$  kpc, and their concentrations  $\{17.3, 15.0, 11.0\}$ , respectively.

electron density is given by:

$$n_e^{hc}(\xi) = n_e^c \left[ 1 + \frac{3.7}{\xi} \ln(1 + \xi) - \frac{3.7}{\xi_c} \ln(1 + \xi_c) \right]^{3/2}, \quad (5.8)$$

where  $\xi_c$  is the dimensionless cooling radius. We will refer to this, more realistic, galactic atmosphere model as “multi-phase”.

### 5.2.3 Galactic atmosphere: kinematics

The second ingredient needed is the velocity field of the free electrons. As for the free electron density, we consider two models: an upper bound and a more realistic rotational velocity. In both cases, we assume the velocity field has cylindrical symmetry.

The upper bound model presumes baryons rotate at the host halo’s circular velocity, which

can be expressed as a function of the cylindrical radial coordinate normalized by the halo's scale radius,  $\varrho$ :

$$v_c(\varrho) = \sqrt{\frac{4\pi G r_s^2 [\ln(1 + \varrho) - \varrho(1 + \varrho)^{-1}]}{\varrho}}. \quad (5.9)$$

The angular momentum of galaxies and their halos is typically expressed in terms of the ratio between the system's angular velocity and the one corresponding to full rotational support, or spin parameter,  $\lambda$ . While for DM halos  $\lambda$  is generally small ( $\lambda \approx 0.05$ , see [234]), for gas it can reach order unity when it collapses towards the halo's center as it cools and is observed in disk galaxies. We refer to a model with  $\lambda = 1$ , whose circular velocity is given by Eq. 5.9 as "fast", or a "fast rotator".

We also consider a more realistic model with  $\lambda < 1$ , and define its velocity field as a fraction of the circular velocity:  $v(\varrho) = f(\varrho, M_v)v_c(\varrho)$ . This fraction depends on the halo's mass and the distance to its center. We use measurements of the tangential velocity of hot gas in hydrodynamical simulations (see Fig. 3 in [235]) to determine  $f$ . For low-mass halos ( $\lesssim 10^{13} M_\odot$ ), this velocity drops from  $\approx 75\%$  of the virial velocity (defined as the circular velocity at the virial radius) in the inner regions to  $\approx 10\%$  at the virial radius. For high-mass halos ( $\gtrsim 10^{13} M_\odot$ ), the ratio of velocities remains roughly constant at  $\approx 10\%$ . Instead of the virial velocity, we use the circular velocity as a normalization, to avoid  $\lambda > 1$  in the halos' innermost regions. We use as a fitting formula:

$$f = \min \left\{ m \log \frac{R}{R_{\text{vir}}} + 0.1, 1 \right\}, \quad (5.10)$$

$$m = -6.0 \times 10^{-6} (\log M_v)^2 + 3.2 \times 10^{-1} \log M_v - 4.4. \quad (5.11)$$

The resulting velocity profile, which we will refer to as "slow", is shown in Fig 5.2, for three different halo masses, together with the alternative, "fast" profiles. As the figure shows, the "slow" profiles are much less sensitive to the halo's mass. The velocities derived from the slow model are in agreement with those predicted for Milky Way and M31 analogs (see [236]).

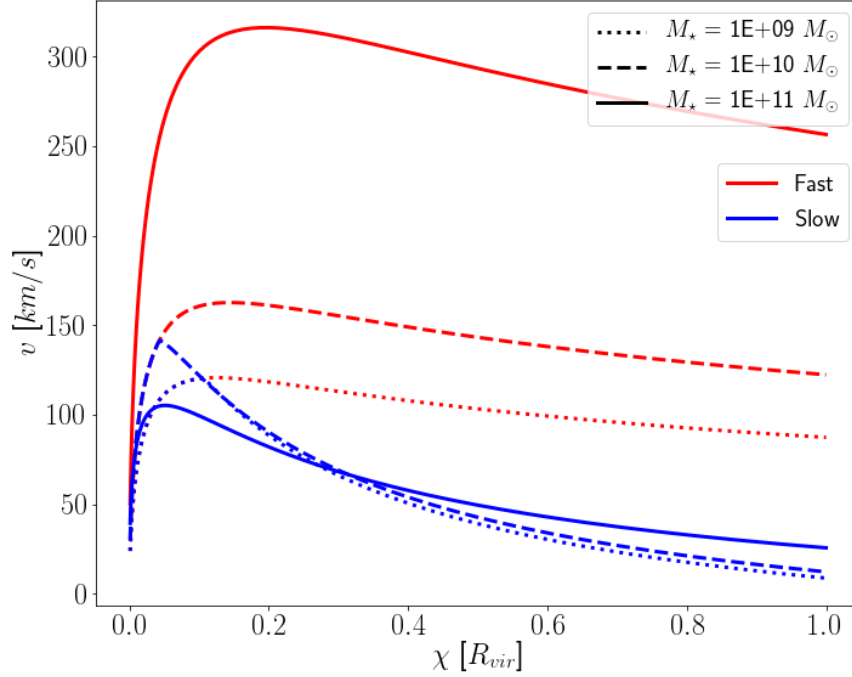


Figure 5.2: Radial profiles of the tangential gas velocity for a fast rotator (in red, rotating at the halo’s circular velocity) and a slow rotator (in blue, rotating at a velocity consistent with simulations [235]) for galaxies of three different stellar masses.

### 5.3 Characterizing the observed rkSZ signal

Detecting the dipole-like kSZ signal induced by the rotation of galactic halos is challenging, compared to that from galaxy clusters. The signal is diminished by the smaller projected electron number density (due to smaller halos), the lower ionization fraction (due to some of the gas cooling and recombining), and smaller angular size on the sky (the beam width for a given CMB experiment will smooth the signal). Stacking the signal from many galaxies is then a necessity.

Different spatial filters can be used to extract the signal from the noise in the stacked data. In this study, we consider two filters: an aperture filter that measures the temperature asymmetry between its right and left halves, and a matched filter designed based on the profile of the expected signal. We next discuss these filters, as well as the expected noise levels and resulting signal-to-noise ratios in both cases.

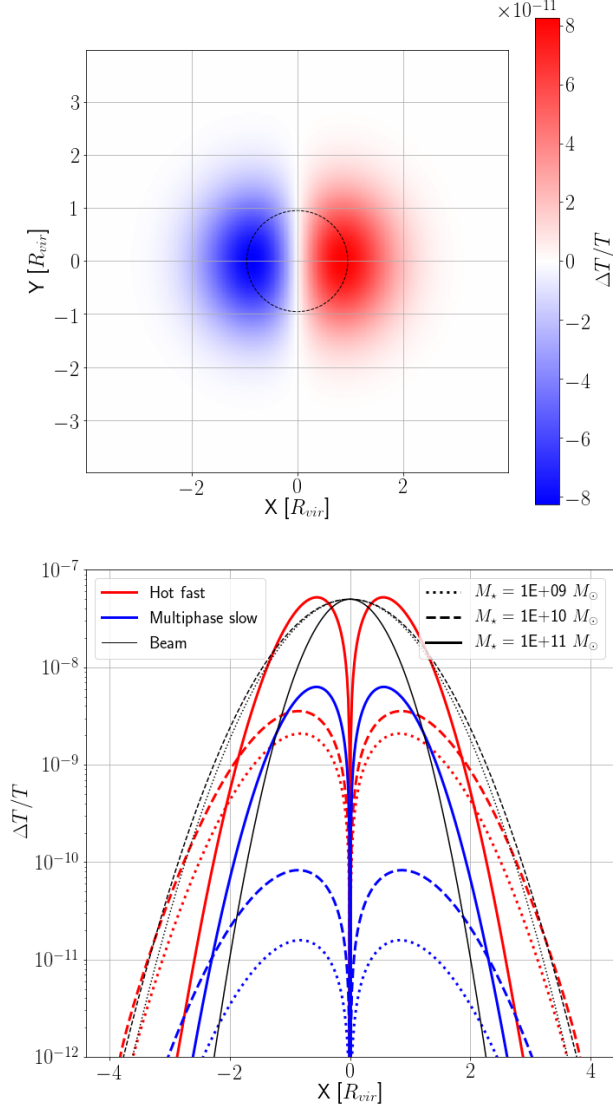


Figure 5.3: **Upper panel:** 2D map of the expected fractional temperature change induced in the CMB by the rotation of a  $M_* = 10^{10} M_\odot$  galaxy hosted by a  $3.4 \times 10^{11} M_\odot$  DM halo with a virial radius of 98.1 kpc and a concentration parameter of  $C_v = 15.0$  at a redshift of  $z = 0.03$ , assuming the multi-phase slow model. The galaxy's inclination is 1 rad and the signal has been convolved with a 5 arcmin FWHM beam (represented by the small dotted circles at the center), while the halo virial radius has an apparent size of 5.2 arcmin. **Lower panel:** cut along the X-axis of the dipole-like signal in the upper panel, for galaxies of three different stellar masses. The predictions in the hot+fast model (fully ionized atmosphere rotating at the halo's circular velocity) are shown in red, and the multi-phase, slow model are shown in blue. The signal was calculated for a metallicity of  $Z_g = 0.3 Z_\odot$ , and shown in absolute value. The profile of the convolving beam is displayed in black.

### 5.3.1 Aperture filter

An aperture filter that measures the temperature difference between its two halves can be used to measure a rkSZ signal, as long as it is centered on the galaxies' and its halves aligned with the galaxies' projected spin vector. A simple statistic is the mean temperature on the right minus the left half of the filter (or dipole):

$$s \equiv \overline{\Delta T}^R - \overline{\Delta T}^L. \quad (5.12)$$

In the absence of a rkSZ effect, we expect this statistic to average to zero. It is a robust statistic, in the sense that it is sensitive to any CMB temperature asymmetry relative to the galaxies' projected spin vectors, regardless of the specific shape of the asymmetry. It is also insensitive to any symmetric (on average) signal induced by the halos, such as the thermal SZ effect (tSZ) or the kSZ effect due to the galaxies' peculiar velocities.

Even in the absence of any kSZ effect, a dipole may arise due to random anisotropies in the CMB within the aperture filter. While the mean dipole due to the CMB's random fluctuations is zero ( $\langle s \rangle_{\text{cmb}} = 0$ ), its variance is not, and should be accounted for as noise. The variance is sourced by both CMB temperature anisotropies and by instrumental noise. Combining both contributions in a single angular power spectrum  $C_\ell = C_\ell^{\text{cmb}} + C_\ell^{\text{noise}}$ , the variance of  $s$  can be computed as (see [237] and Appendix D):

$$\sigma_s^2 = \left\langle \left( \overline{\Delta T}^R \right)^2 \right\rangle + \left\langle \left( \overline{\Delta T}^L \right)^2 \right\rangle - 2 \left\langle \overline{\Delta T}^R \overline{\Delta T}^L \right\rangle = 2 \left[ \left\langle \left( \overline{\Delta T}^R \right)^2 \right\rangle - \left\langle \overline{\Delta T}^R \overline{\Delta T}^L \right\rangle \right] \quad (5.13)$$

The aperture filter is defined by its window function  $W(x, y)$  in a coordinate system in which the galaxies' spin is aligned with the  $y$ -axis. The covariance between the mean temperatures measured over the two halves of the window function by an instrument whose beam profile in Fourier space, or beam function, is  $b_\ell$ , can be estimated by

$$\left\langle \overline{\Delta T}^R \overline{\Delta T}^L \right\rangle = \int \frac{d^2 \ell}{(2\pi)^2} b_\ell^2 C_\ell \tilde{W}_L^*(\ell) \tilde{W}_R(\ell), \quad (5.14)$$



where  $\widetilde{W}^*$  is the Fourier transform of  $W(x, y)$ . The variance in each of the two halves follows from the analogous expressions, but with  $|\widetilde{W}_L|^2$  or  $|\widetilde{W}_R|^2$  in the integrand. The window function for the aperture filter is semi-analytic in Fourier space (see Eq.D.3).

### 5.3.2 Matched filter

The aperture filter described in § 5.3.1 is robust, but not optimal, since not all the information encoded in the shape of the signal is used. The optimal approach would be to use a matched filter (see, for example, [238]). For each galaxy, the optimal filter is essentially the expected rkSZ signal pattern, with the different angular scales weighted by the expected noise (CMB anisotropies and instrumental noise). In Fourier space,

$$\widetilde{\text{MF}}(\ell) = \frac{1}{\int d^2\ell \frac{|\widetilde{\Delta T}_{kSZ}(\ell)|^2}{C_\ell}} \frac{\widetilde{\Delta T}_{kSZ}^*(\ell)}{C_\ell}. \quad (5.15)$$

For each galaxy, this filter can be applied to the corresponding CMB data, and the result stacked for all galaxies in the survey. The expected signal will make itself apparent as a high peak at the center of the stack. In the absence of any signal, the filtered data will yield pure noise. The height of the central peak relative to the standard deviation in the absence of signal can be used as an estimate for the signal-to-noise ratio (SNR) of the detection via this approach.

Note that additionally, a matched filter is optimal only if the model used for its design corresponds to the true signal in the data. Also, contrary to the aperture filter, the matched filter is not insensitive to potential isotropic signals induced by galactic halos, such as the tSZ effect or kSZ effect induced by peculiar velocities (see § 5.6.1).

## 5.4 Measurement signal-to-noise and required number of galaxies

In order to coherently stack CMB data for each galaxy, without nulling their rkSZ signal, we need to align the CMB data with each galaxy's projected spin angle. These angles can be measured, for instance, from spatially-resolved spectroscopic data. Integral field spectroscopy

enables the efficient acquisition of such data for thousands of galaxies. Examples of recent and ongoing surveys include MaNGA <sup>4</sup> and SAMI <sup>5</sup>.

To assess the viability of measuring the stacked rkSZ signal, we estimated the number of galaxies needed for a  $3\sigma$  detection.

We considered galaxy surveys with the same redshift and stellar-mass distributions as MaNGA (specifically its "primary" sample, see § 5.5.1) and SAMI. We divided each survey's range of redshift and stellar mass in a  $40 \times 40$  grid, resulting in 235 non-empty bins for MaNGA and 358 for SAMI, shown in Figure 5.4. The expected signal,  $s_i$ , and noise,  $\sigma_i$ , contributed by each bin is estimated from the mean redshift  $\langle z_i \rangle$  and stellar mass,  $\langle M_{\star i} \rangle$  of galaxies in each bin, for simplicity.

Assuming galaxies are randomly oriented, the probability density function of their inclination angle  $i$  is  $p(i) = \sin i$ , and the mean inclination is  $\langle i \rangle = 1$  rad, which is the value we adopted for all bins. The SNR for the full survey is a weighted average of the signal and noise in each bin:

$$\text{SNR} = \sqrt{N_{\text{gal}}} \frac{\sum_{i=1}^{N_{\text{bin}}} w_i f_i s_i}{\sqrt{\sum_{i=1}^{N_{\text{bin}}} w_i^2 f_i \sigma_i^2}} \equiv \sqrt{N_{\text{gal}}} \text{SNR}_1. \quad (5.16)$$

Here  $N_{\text{gal}}$  is the total number of galaxies in the survey, distributed among  $N_{\text{bin}}$  bins, each with a fraction of the total  $f_i$ , and  $\text{SNR}_1$  is the equivalent single-galaxy SNR, which depends on the average properties of the survey's galaxies. After some algebra, it can be shown that the weights that maximize the survey's SNR are  $w_i = s_i / \sigma_i^2$ .

Future high-resolution CMB experiments will be able to resolve the halos of nearby galaxies. For relatively large scales ( $\ell < 10^3$ ), primordial CMB fluctuations are the dominant source of noise. At smaller scales, we also consider instrumental noise for the five different experimental configurations listed in Table 5.1. Each experiment is characterized by its beam's FWHM and its instrumental noise, which is defined by a white,  $\ell$ -independent power spectrum [239].

The first configuration in Table 5.1 corresponds to *Planck* (we use as a reference its 217 GHz

---

<sup>4</sup>Mapping Nearby Galaxies at APO: <https://www.sdss.org/surveys/manga/>

<sup>5</sup>Sydney-Australian-Astronomical-Observatory Multi-object Integral-Field Spectrograph: <https://sami-survey.org/>

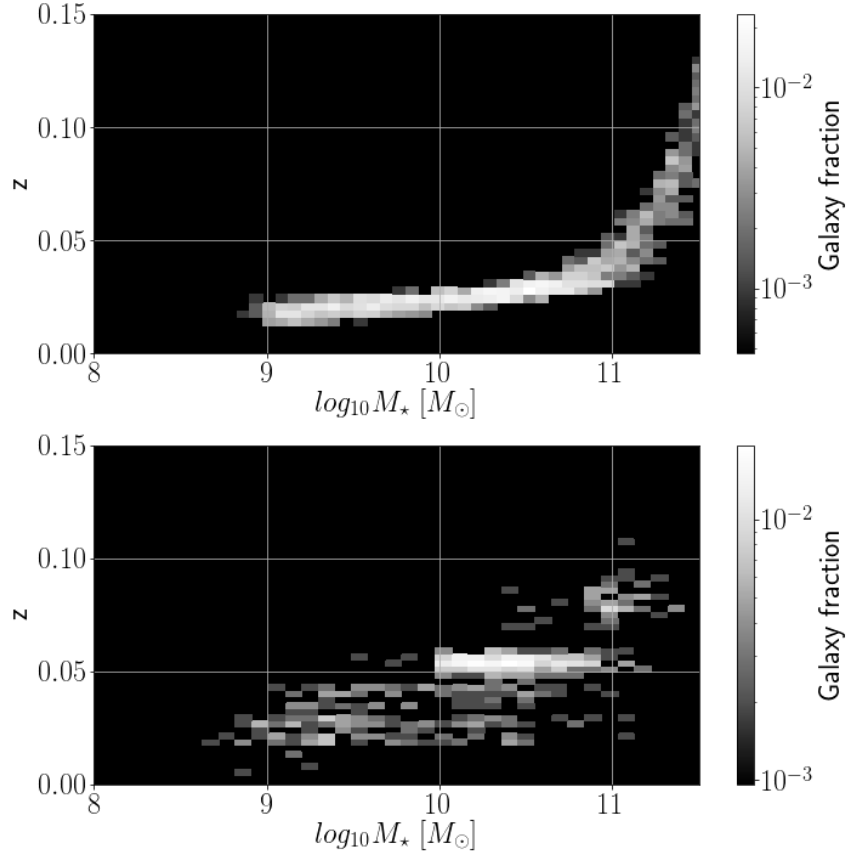


Figure 5.4: The distribution of galaxies in redshift and stellar mass, in the two prototype surveys considered to assess the feasibility of detecting the kSZ signal induced by the rotation of galactic halos. The top panel corresponds to the primary sample in MaNGA, and the lower panel to SAMI.

Experiment	FWHM [arcmin]	$\Delta T_{\text{noise}}$ [ $\mu K$ arcmin]
<i>Planck</i>	5.00	45.6
ACT	1.40	15.0
Simons	1.40	6.0
CMB-S4	1.40	1.0
CMB-HD	0.25	0.5

Table 5.1: Instrumental configurations considered for different existing and planned CMB experiments, defined in each case by the beam’s FWHM and the rms noise level.

channel, whose frequency is close to the one at which the tSZ is null). The second is the 148 GHz channel from ACT (the Atacama Cosmology Telescope <sup>6</sup>), the third is the “Goal” target for the 145 GHz channel of the Simons observatory (Simons observatory <sup>7</sup>), the fourth is a possible high-frequency channel of a CMB stage 4 experiment (CMB-S4 <sup>8</sup>) and the fifth a potential high-resolution future CMB experiment, (CMB-HD [240]).

For illustration, in Figure 5.5 we show a comparison of the power spectrum of the intrinsic temperature anisotropies of the CMB (computed with CAMB [194]) and of the instrumental noise for all five configurations.

Also shown in the figure, for reference, is the power spectrum of the rkSZ signal,  $|\Delta T_{kSZ}(\ell)|^2$ , for galaxies with three different stellar masses (defined simply as the 2D Fourier transform of the signal in Eq. 5.1 and shown in Fig. 5.3). The labels on the y axis on the left correspond to the CMB and the instrumental noise, and on the right to the kSZ signal. The large difference in magnitudes is indicative of the large number of galaxies that will need to be stacked to required to separate the signal from the noise.

We used the two filters described in § 5.3 to compute the signal and the noise contributed by each galaxy bin. The measured signal results from applying the filters to the expected theoretical kSZ signal from the two models detailed in § 5.2: a hot, fast-rotating and a multi-phase, slow-rotating galactic atmosphere. The signal for the aperture filter is the magnitude of the measured temperature dipole, given in Eq. 5.12, whereas for the matched filter, it is the height of the central

<sup>6</sup>Atacama Cosmology Telescope:<https://act.princeton.edu/>

<sup>7</sup>Simons Observatory:<https://simonsobservatory.org/>

<sup>8</sup>CMB-S4:<https://cmb-s4.org/>

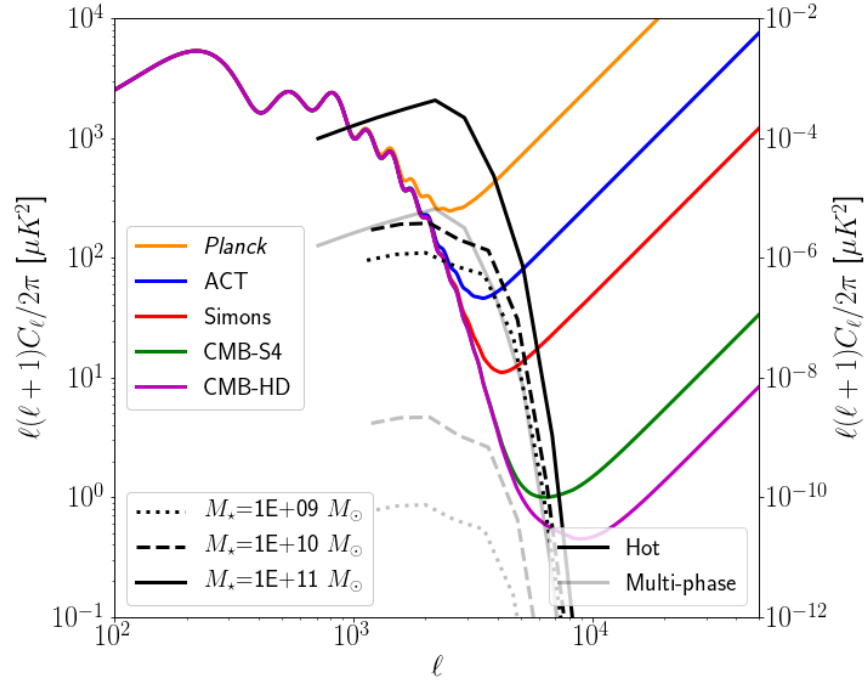


Figure 5.5: Comparison between the power spectrum of CMB temperature anisotropies, including instrumental noise which dominates at  $\ell \gtrsim \text{few} \times 10^3$ , and that for the expected rkSZ signal for three galaxies of different stellar mass (the same galaxies used in Figs. 5.1-5.3). The scale on the left y-axis refers to the CMB+noise and the scale on the right to the kSZ power spectra. Note that the rkSZ power is several orders of magnitude lower than that from the combination of CMB + instrumental noise.

peak of its convolution with the expected theoretical signal, as discussed in § 5.3.2.

The noise level for the aperture filter is computed directly from Eq. (5.13), using the numerical  $C_\ell$  that includes both the primary CMB and the instrumental noise (shown in Fig. 5.5). For the matched filter, in each redshift and stellar-mass bin  $i$ , we created 100 independent realizations of synthetic noise-only CMB maps. Each synthetic noise map is generated from a Gaussian random field, defined again by the combined power spectrum of the CMB and the experiment’s instrumental noise. Each map is then convolved with the matched filter for the mean redshift  $\langle z_i \rangle$  and stellar mass,  $\langle M_{\star i} \rangle$  in that bin and yields a peak height in real space; the noise is computed as the standard deviation of these 100 peak-height values. This exercise yields the signal-to-noise ratio per galaxy  $\text{SNR}_1$  in Eq. (5.16), and the number of galaxies  $N_{\text{gal}}$  required for a  $3\sigma$  detection follows by setting the total  $\text{SNR}=3$  in this equation.

Fig. 5.6 shows the number of galaxies required for a  $3\sigma$  detection using the aperture filter, as a function of the filter’s size, for the set of CMB experiments under consideration. It assumes a MaNGA-like survey (the results for a SAMI-like survey are qualitatively the same with slightly higher number requirements). For a given experiment, the number of galaxies needed decreases as the size of the aperture filter gets smaller, up to the point where the aperture can resolve the galaxies’ halos. The resolution of CMB experiments is the main factor that determines their ability to measure the rkSZ signal, over the level of instrumental noise.

A summary of the number of galaxies needed, for both surveys, is shown for each CMB experiment configuration in Table 5.2. The reference aperture size is  $0.1 R_{\text{vir}}$ . As Fig. 5.6 shows, the number of galaxies required decreases dramatically as the resolution of the CMB experiments improves. The higher redshift of SAMI galaxies translates into a somewhat larger number of galaxies needed for a detection. The table also shows that matched filtering can reduce the number of galaxies needed, compared with the aperture filter, by a factor of  $\approx 2$  for *Planck*, and by more than an order of magnitude for a CMB Stage 4 experiment. However, this statistic is more sensitive to modeling errors, filter misplacements, and isotropic signals on the scale of galactic halos (tSZ, kSZ from peculiar velocities).

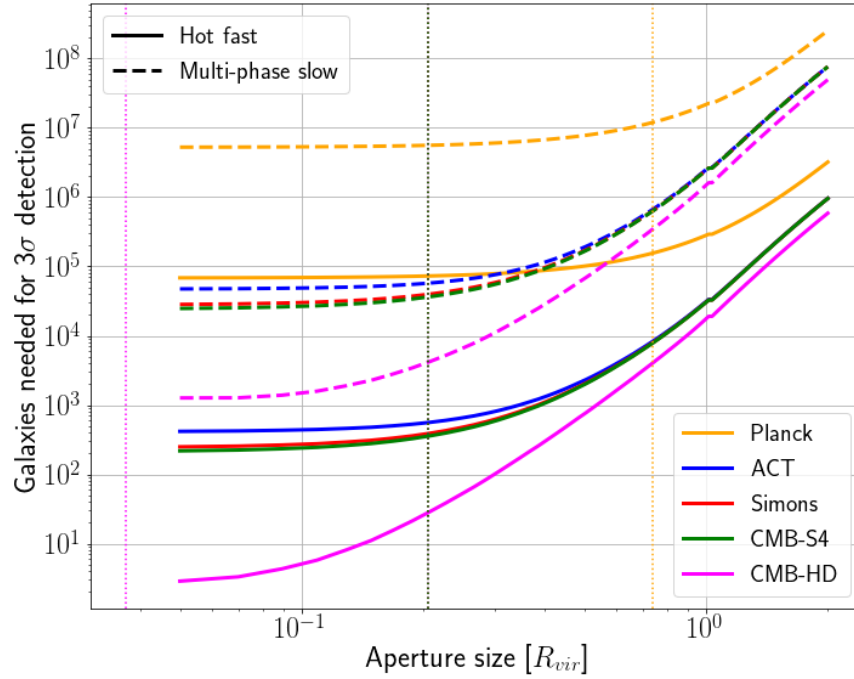


Figure 5.6: Number of galaxies needed for a  $3\sigma$  detection of the rkSZ signal using an aperture filter, as a function of filter size. Each CMB experiment configuration is displayed in a different color. The galaxy  $(z, M_\star)$  distribution is assumed to be that of the primary sample of MaNGA. Solid lines correspond to predictions based on the hot+fast model, while dashed lines are based on the multi-phase, slow model. For reference, vertical lines represent the CMB experiment beam size, in units of the average  $R_{vir}$  for the galaxies in the survey.

	MaNGA-like	SAMI-like
	Hot, fast rotating corona	
<i>Planck</i>	6.9e4   2.2e4	5.9e4   2.7e4
ACT	4.4e2   3.5e2	1.6e3   1.0e3
Simons	2.8e2   6.7e1	1.1e3   2.2e2
CMB-S4	2.5e2   4.0e0	9.4e2   1.4e1
CMB-HD	6.0e0   1.0e0	1.2e1   2.0e0
	multi-phase, slow rotating corona	
<i>Planck</i>	5.3e6   1.6e6	<u>7.5e6</u>   3.5e6
ACT	4.9e4   3.1e4	2.3e5   1.4e5
Simons	3.1e4   6.1e3	1.4e5   2.9e4
CMB-S4	2.7e4   4.1e2	1.3e5   2.0e3
CMB-HD	1.6e3   7.8e1	7.1e3   4.0e2

Table 5.2: The number of galaxies required for a  $3\sigma$  detection of a rkSZ signal. For each combination of CMB experiment and galaxy survey type, the number on the left corresponds to the aperture filter (measured at  $0.1 R_{\text{vir}}$ ) and the one on the right to matched filtering. The combination of *Planck* and a SAMI-like survey, using the aperture filter and adopting the multi-phase slow-rotator model, yields a required number of galaxies exceeding those available within the stellar-mass and redshift range of the galaxy survey. This detection is therefore impossible, and is marked in italics (see § 5.6.3 for discussion).

While the numbers are large for all the cases, future CMB experiments will be able to rule out most models and may be able to make a detection (see further discussion of the detection feasibility in § 5.6.3 below).

## 5.5 Stacking *Planck* data at the positions of MaNGA galaxies

As a proof-of-concept, we stacked *Planck* data at the positions of galaxies from the MaNGA survey. While the number of galaxies is insufficient to make a rkSZ detection, it can yield an upper limit on the average CMB temperature dipole aligned with galaxies’ spin.

### 5.5.1 Galaxy data: MaNGA

MaNGA is an integral field survey with the goal to acquire spatially-resolved spectroscopy from  $\approx 10,000$  galaxies [241]. Galaxies were targeted to follow a (roughly) flat distribution with respect to their stellar mass, along two different sequences [242]. The first, or “Primary” sample, consists of low-redshift galaxies for which MaNGA’s IFU spectrographs cover  $\approx 1.5$  times their



half-light radius,  $R_e$ . The second, or “Secondary” sample, is comprised of higher-redshift galaxies, for which MaNGA’s spatial coverage increases to  $\approx 2.5 R_e$ . A third sample, the “Color-enhanced supplement”, increases the survey’s galaxy count in areas of the Primary sample’s color-magnitude diagram that are otherwise poorly sampled. Galaxies in the Primary sample contribute the most to the overall SNR of the stacked kSZ signal, because its galaxies have larger apparent size on the sky (for a given stellar mass), and therefore their kSZ signal is suppressed less by the CMB beam’s smoothing.

Among other data products, MaNGA provides, for each observed galaxy, two-dimensional maps of line-of-sight velocities separately for stars and gas. This kinematic information can be used to estimate the galaxy’s spin angle projected on the sky, hence the usefulness of this survey to try to detect any effect (on average) of the galaxies’ rotation.

After applying a series of quality cuts described in Appendix E to MaNGA’s DRPa11 catalog, made publicly available as part of SDSS’s data release DR15, our stacking sample consists of 2,664 galaxies: 1,231 are part of the Primary sample, 982 are part of the Secondary sample, and 451 of the Color-enhanced supplement.

We used additional value-added catalogs to access information about the galaxies in our stacking sample that is not included in the DRPa11 catalog. The MaNGA Morphology Deep Learning DR15 catalog supplies information on the galaxies’ morphology. The morphological classification is performed using an automated model trained and tested on SDSS-DR7 images [243]. The 2,664 galaxies are split by type as follows: 282 are ellipticals, 424 are S0s, 1,953 are spirals, and 5 are classified as irregulars.

Finally, the Galaxy Environment for MaNGA Value Added Catalog (GEMA-VAC, [244]) provides environmental information based on the sign of the eigenvalues of the tidal tensor at the location of each galaxy [245]. In our stacking sample, 1,056 galaxies are in a cluster environment, 1,239 in filaments, 331 in sheets, and 38 in voids.

### 5.5.2 CMB data: *Planck*

We used publicly available CMB maps from *Planck* [246], specifically the full mission, temperature SMICA-noSZ map and full mission, single-frequency maps from *Planck*'s high frequency instrument. The SMICA-noSZ map is a linear combination of multi-frequency CMB maps that cancels any contribution with the spectrum of the tSZ effect, leaving the kSZ signal, which has the same spectrum as the CMB, unaffected. It also cleans other foreground signals, based on their contribution to the variance in the data.

On average, the tSZ signal from galaxy halos should be isotropic. While measurements with the aperture filter described in § 5.3.1 are insensitive to isotropic signals, measurements based on a matched filtering could be affected, as we discuss in § 5.6.1. It is therefore preferable to use data that has already been cleaned from other halo-induced temperature anisotropies, such as the tSZ. The SMICA-noSZ map has a HEALPix resolution  $N_{pix} = 2048$  and a spatial angular resolution of 5 arcmin FWHM. We combined the temperature data with the common temperature confidence mask before performing any measurement, and verified that our results do not depend on the mask used.

### 5.5.3 Stacking

CMB data needs to be aligned with the galaxies' projected spin angle. Otherwise, any possible rkSZ signal would be cancelled. We also scaled the *Planck* data with the angular diameter of each galaxy's halo, to add the signal profiles coherently.

Since we are interested in capturing the rotation of ionized gaseous halos, we estimated the spin angle using the emission line with the shortest wavelength (highest photon energy) for which MaNGA provides kinematic information: O II. For both of the two lines that form the doublet (3,727 Å and 3,729 Å, corresponding to temperatures of  $3.8 \times 10^4$  K), we computed the momentum

of their line-of-sight velocity map relative to the galaxy’s position in the catalog:

$$\mathbf{L} = \sum_i F_i \mathbf{v}_i \times \mathbf{r}_i \quad (5.17)$$

For each spaxel,  $i$ , the line-of-sight velocity,  $\mathbf{v}_i$ , and the flux,  $F_i$ , are weighted with their inverse variance before entering the cross-product with the spaxel’s position vector,  $\mathbf{r}_i$ . The projected spin angle for each line is the angle between  $\mathbf{L}$  and the North direction. The spin angle for the galaxy is the average of the spin angle for each of the doublet’s lines.

We confirmed that our results do not depend on the specific line used to estimate the spin angle. We reached the same results using Ne III, and angles estimated from all emission lines measured by MaNGA are highly correlated with each other. This is not surprising, since MaNGA probes the inner regions of halos (up to  $\approx 1.5 R_e$  or  $\approx 2.5 R_e$ ), where gas kinematics tend to be coherent. One of the criteria used to select the galaxies used for this analysis was precisely that the spin angle did not depend strongly on the tracer used to compute it; we rejected galaxies for which the spin angle estimated from O II differed significantly from the one estimated from H $\alpha$  (see Appendix E). A visual inspection showed that most of these rejected galaxies have complex kinematics or are face-on systems that would not contribute to the rkSZ signal.

MaNGA measures the gas kinematics only in the inner regions of galaxies, on average up to  $\approx 2\%$  of the galaxies’ virial radius ( $\approx 5 - 6\%$  in the most highly resolved cases). By comparison, the rkSZ signal peaks further out, in the inner halo, at  $0.1 - 0.2 R_{\text{vir}}$  (depending on the CMB experiment’s beam size). This corresponds to an extrapolation by a factor of several in spatial scale; a key assumption in our analysis is therefore that the galaxies’ gaseous halos at this radius co-rotates with their inner region where the spin is measured. In a hierarchical formation scenario, this is not necessarily the case, for outer halos may have built up from contributions with different angular momenta [247]. Nevertheless, recent measurements on simulations show a strong correlation in the angular momentum of hot gas across most of the virial radius [248]. A non-detection can then be due to the lack of correlation between the rotations of the inner and outer regions of gaseous halos. However, in order to significantly suppress the rkSZ signal, the r.m.s. variation in

the estimated spin angle would have to be large (see § 5.6.2 and Appendix C).

We scaled the CMB data with the angular diameter of each galaxy’s host DM halo. This angle is fully determined by the galaxies’ redshifts and their stellar masses, both found in MaNGA’s DRPall catalog. Stellar masses, which are inferred from the galaxies’ Sersic photometry (hence their  $h^{-2}$  cosmological dependency), were converted to DM halo masses and virial radii (see § 5.2).

Matched filtering requires a template for the expected signal. We built two templates for each galaxy, based on the hot fast-rotator and multi-phase slow-rotator models described in § 5.2. A parameter that needs to be derived from the data to generate the templates is the galaxies’ inclination angle,  $i$ . To do so, we modeled galaxies as oblate spheroids, for which [249]

$$\cos i = \sqrt{\frac{\frac{b}{a} - q^2}{1 - q^2}}. \quad (5.18)$$

We used the Sersic axial ratio,  $b/a$ , from the DRPall catalog, and assumed that galaxies have aspect ratios of  $q = 0.15$  when seen edge-on. Furthermore, we assigned an inclination of  $90^\circ$  to all galaxies whose edge-on probability reported in the morphology catalog exceeds 99%. This estimation is meaningless for elliptical galaxies.

Another reason why elliptical galaxies are problematic for our stacking analysis is that they have, in general, small spin parameters. For these reasons, in addition to the stack analysis of the 2,664 galaxies selected from MaNGA’s DRPall catalog, we also performed an analysis restricted to the 1,953 spiral galaxies in the sample. Furthermore, the environment can affect the properties of galactic gaseous halos. The kinematics of the outer halos of spiral galaxies in clusters may be perturbed by close encounters with other cluster members and interactions with the intra-cluster medium. Thus, we performed a third analysis restricted to spiral galaxies which do not reside in a cluster environment (field spirals, 1,235 galaxies).

The results of stacking *Planck*’s CMB data on the positions of these three galaxy samples are displayed in Fig. 5.7, for the case in which all galaxies are equally weighted. The results of the analyses do not change when optimal weights based on the expected signal are used. Visually, the

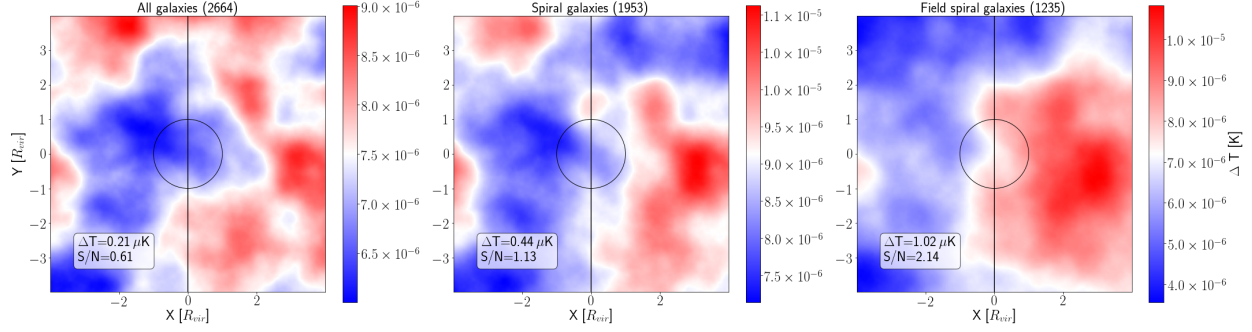


Figure 5.7: *Planck* SMICA-noSZ CMB data stacked on the positions of MaNGA galaxies (with equal weight) after aligning with the galaxies’ spin angles, and scaling to their  $R_{\text{vir}}$ . The left panel shows the results for all 2,664 galaxies in our sample, the central panel for the 1,953 spiral galaxies, and the right panel for the 1,235 field spiral galaxies. For each galaxy, located at the origin (0,0), we stacked spin-aligned  $8 \times 8 R_{\text{vir}}$  patches. The circles correspond to  $R_{\text{vir}}$  and the vertical lines mark the expected galaxy spin direction (pointing towards the -y axis, the right half is approaching the observer, and the left side receding). Any rotation-induced temperature dipole should show a left-right cold-hot temperature asymmetry (see Fig. 5.3). In each text box, the measured dipole on a  $1.0 R_{\text{vir}}$  aperture, and its signal-to-noise ratio based on the theoretical noise calculation described in 5.3.1.

stack corresponding to field spirals (right panel) shows a temperature asymmetry with the correct sign if it were induced by the rotation of the galaxies (the spin angle points towards the negative Y axis). Its amplitude — $1.02 \mu K$  measured on a filter with an aperture of  $1.0 R_{\text{vir}}$ — is large compared with the expected signal —a mere  $2.7 \times 10^{-2} \mu K$  if the hot, fast rotator is a good model—, but still only  $2.14\times$  larger than the expected noise for that aperture. The other two samples (middle and left panels) appear consistent with noise.

An alternative way to measure the significance of any dipole measured on stacked data is to draw a large number of measurements on noise-only data, after randomizing the positions and orientations of the filter. We show the result for  $10^4$  such measurements in Fig. 5.8. The noise-only map on which the measurements were done is the result of stacking the CMB data on the positions of the galaxies assuming a random orientation of their spins. The measurements follow a Gaussian distribution, and the measured dipole of  $1.02 \mu K$  corresponds to a 94.65 percentile, or a significance of  $1.63\sigma$ . This is lower than the  $2.14\sigma$  significance derived from the theoretical calculation of the noise for that aperture. While not significant enough to be a detection, the measured dipole is suggestive enough to raise the question of whether it is real, and if so, what

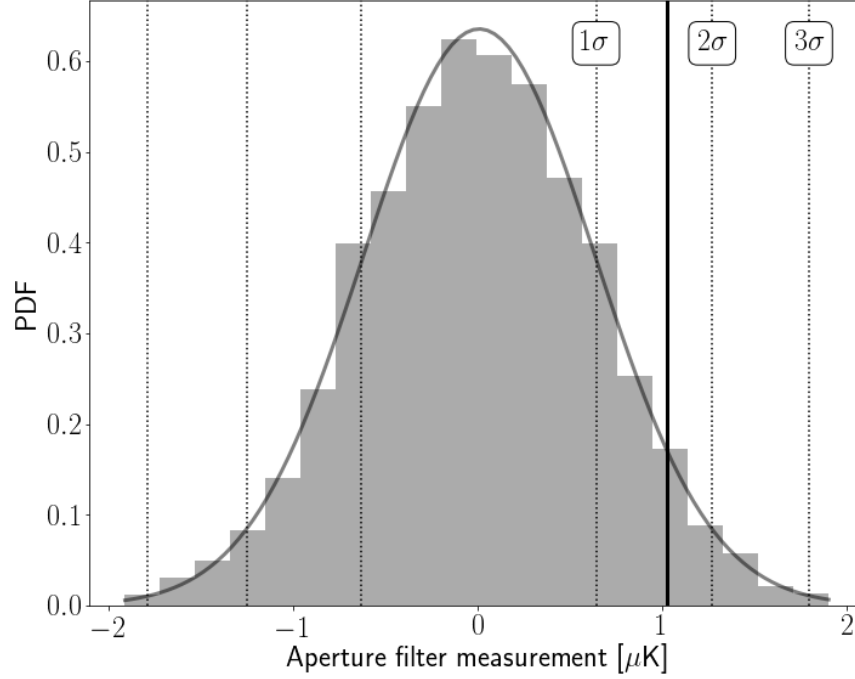


Figure 5.8: Probability distribution function (PDF) of the temperature dipole of the SMICA-noSZ map stacked on the positions of 1,253 field spirals after randomizing their spin angles. The PDF is inferred from  $10^4$  measurements using  $1.0 R_{\text{vir}}$  aperture filters that have been randomly placed and rotated. Superimposed, a Gaussian PDF with the same mean and standard deviation as the  $10^4$  measurements, shows good agreement with the data. Dotted vertical lines correspond to  $1\sigma$ ,  $2\sigma$ , and  $3\sigma$  thresholds. The solid vertical line is the measured dipole on the stack with the galaxies' spins aligned.

could cause such an unexpectedly large signal. To answer that question, we performed the same analysis on *Planck*'s single frequency maps (see Fig. 5.9).

In the 100, 143, and 217 GHz data, the stacks look similar to the one from the SMICA-noSZ map. The significance of the measured dipoles, estimated from  $10^4$  measurements on noise-only stacks (stacks with randomized galaxy spins), is lower than that of the SMICA-noSZ dipole, at  $1.42\sigma$ ,  $1.80\sigma$ , and  $0.96\sigma$ , respectively. Higher frequency maps, at 353, 545, and 857 GHz, look different, with a clear signal coming from within the galaxies' virial radius, and a small dipole with of opposite sign to the one measured on the SMICA-noSZ map. The lack of a consistent dipole across frequencies, and the low significance of the measurements indicate that there is no real temperature dipole in our data above the noise level. These results are robust to the choice of CMB mask, size of the CMB patches and their weighting for stacking. Finally, matched filtering,

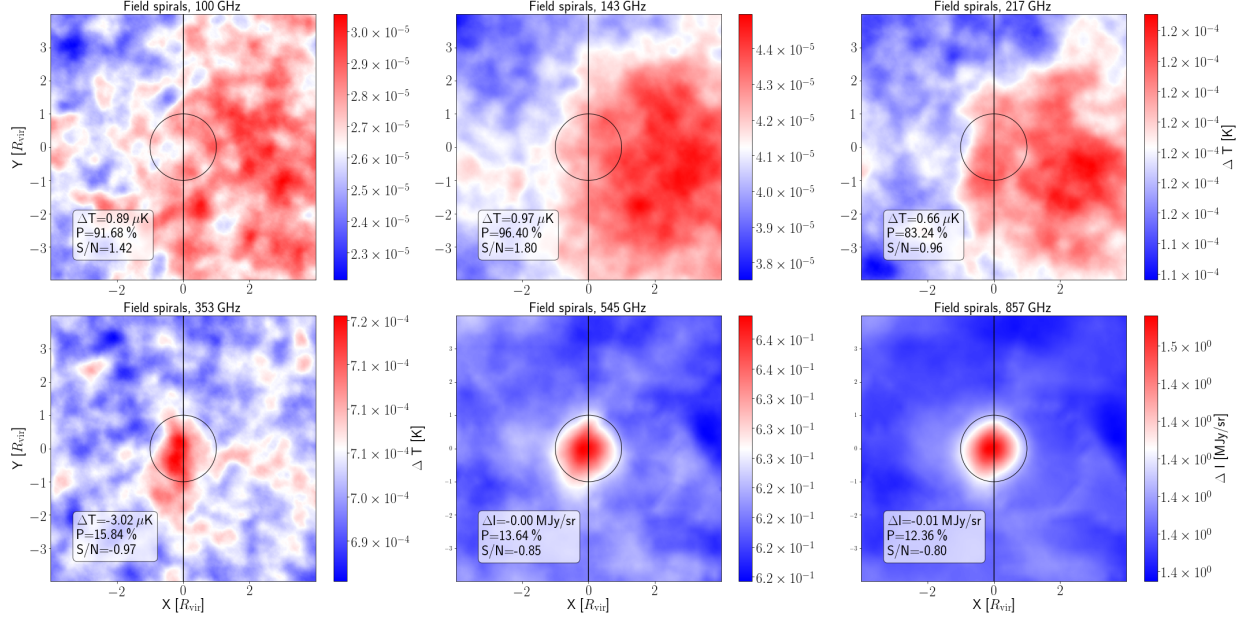


Figure 5.9: Single-frequency *Planck* CMB maps from its high-frequency instrument (HFI), stacked with equal weights on the positions of the 1,235 field spirals after aligning their spin angles. Note the difference in units for the two highest-frequency maps: MJy/sr instead of K. As with Fig. 5.7, the circle represents the average virial radius and the vertical line the spin direction. The text boxes show the measured dipole, as well as its percentile and S/N ratio. The percentile and significance have been derived from noise-only stacks.

using the two models for galactic atmospheres described as part of this work, does not uncover a signal either.

## 5.6 Discussion

### 5.6.1 Model and observational uncertainties

The estimates shown in Table 5.2 do not include model uncertainties. The models used to describe the gaseous component of galactic halos are a simplification. For instance, the density profiles are spherically symmetric and the pressure profiles do not account for any rotational support. For a feasibility study, and given the current uncertainties about the properties of galactic atmospheres, we deem the level of detail of the kSZ models sufficient, and leave for future work the use of more sophisticated models, such as the self-consistent rotating profiles developed in [250], or models taken directly from hydrodynamical simulations.

Two model parameters that could modify significantly the strength of the kSZ signal are the total mass of baryons in the halo, parametrized by the halo baryon fraction, and the galaxy’s peculiar velocity. An effect on halo scales that we have not included in our models is the tSZ.

Our models assume that the halo baryon abundance matches the cosmic average,  $M_b = f_b M_h \equiv (\Omega_b/\Omega_m)M_h$ . In practice, only a small fraction of these baryons are observed in galaxies, which is known as the missing baryon problem. These baryons could reside in the IGM, and the galactic halos could, as a result, be baryon-poor relative to the cosmic abundance. A lower baryon fraction would reduce the electron density in the halo, and therefore its rkSZ signal.

For the limiting case in which all the gas is ionized, the electron density scales linearly with the baryon fraction through the multiplicative factor of  $\rho_0$  in Eq. (5.4). A baryon fraction of half the cosmic abundance would reduce electron densities and the kSZ signal by a factor of two and the number of galaxies needed for a detection would increase by a factor of four. For the more realistic, multi-phase model, the impact of the overall baryon fraction in the halo is more complex. The cooling density in Eq. (5.7) is independent of the total baryonic mass of the halo (as long as DM dominates the gravitational potential and determines the temperature of the hot gas in hydrostatic equilibrium). The cooling radius depends on the baryon fraction, since it is partly determined by the initial distribution of hot gas, and it affects the density profile of the hot corona through  $\xi_c$  in Eq. (5.8). As a result, the residual hot corona in the multi-phase model is less sensitive to the baryon fraction than the gas in the hot model. Physically, in the multi-phase model, the reduction in the baryons is primarily absorbed by the cold gas, except in the outer regions, where the small densities contribute little to the SNR of the rkSZ signal.

Fig. 5.10 shows the effect of a reduced baryon fraction on the free electron density. A reduction in the baryon fraction by a factor of two reduces the electron density in the inner parts of multi-phase halos only by  $\lesssim 10\%$ . Consequently, the number of galaxies needed for a detection may not be as sensitive to the baryon fraction in galactic halos may naively be expected.

When estimating the number of galaxies needed for a detection, we also did not take into consideration the galaxies’ peculiar velocities. Their line-of-sight velocity, relative to the Hubble flow,



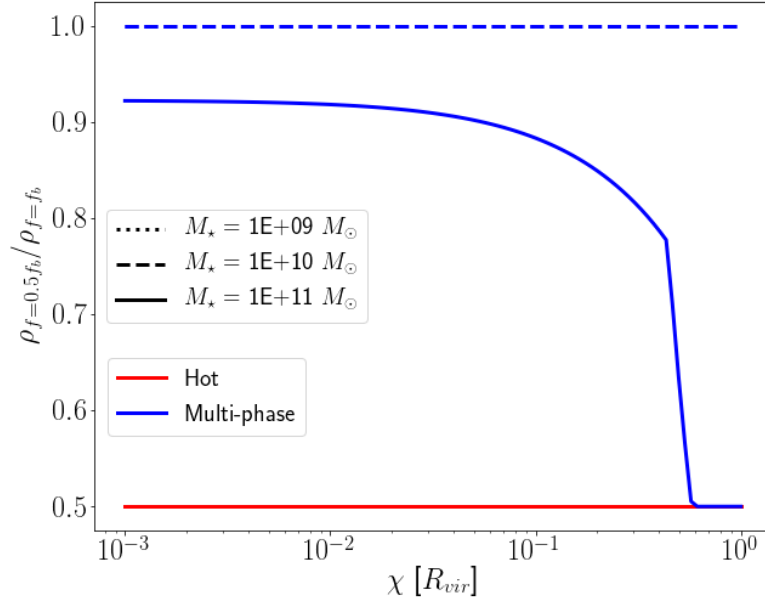


Figure 5.10: The figure illustrates the impact of a factor-of-two reduction in the total baryon mass in the halo relative to the cosmic mean value. The corresponding reduction in the density of the hot, ionized gas component is shown as a function of radius, in units of the virial radius. Results are shown for the same three halos as in the previous figures, labelled by the stellar mass of their central galaxies. In the hot model, where all the baryons are ionized, the gas density scale linearly with the baryon fraction, independent of mass and radius (shown in red). The hot coronae of the multi-phase gaseous halos (shown in blue) are less affected, except in the outer regions where the rkSZ signal contributes little to the SNR. The effect on the multi-phase model decreases with smaller halo mass.

can be significant [251] and the induced kSZ effect dominant, compared to the rkSZ. Nevertheless, since the induced temperature shift is symmetric with respect to the galaxies' spin axes, the aperture filter is blind to this shift. In the absence of centering errors, matched filters are also insensitive to a symmetric signal. The reason is that the matched filter has odd parity relative to the spin axis, and therefore a convolution with a signal that has even parity relative to the same direction leaves it unchanged (including the height of the central peak). In the presence of centering errors (or anisotropies of the kSZ signal induced by the peculiar motions), this conclusion no longer holds. Fortunately, when stacking a large ensemble of galaxies with random peculiar velocities, the effect should still average out.

The tSZ signal from galactic halos is symmetric relative to their spin axes, like the kSZ from peculiar motions, but contrary to the peculiar-motion-induced kSZ, its sign is not random. Subsequently, in the presence of filter centering errors, its effect on matched filtering will not average out, and measurements will be affected. The aperture filter is still insensitive to this effect. When applying matched filtering to CMB data for stacking, either the tSZ signal should be modelled and incorporated in the analysis, or the CMB data should be cleaned from the tSZ signal. This is the reason we used the `SMICA-noSZ` map in § 5.5.

The significance of any future detection should be assessed by considering these factors, together with any observational uncertainties in the parameters that inform the models, such as spin angles, stellar masses, scatter in the stellar mass to halo mass relation, etc. Such an analysis is beyond the scope of the present study but will be warranted if/when a detection is claimed.

### 5.6.2 Measurement uncertainties

Misplacing the filters used to measure the rkSZ effect relative to the signal is a source of errors, independently of the model used to interpret the measurements. A filter misplacement can be a decentering relative to the galaxy, a misalignment of the filter's axis with the galaxy's projected spin vector, or a combination of the two. The average effect of these misplacements (assuming they are random) can be characterized as a suppression of the measured kSZ signal. If the probability

distribution of these positioning and alignment errors,  $\epsilon$ , is given by  $p(\epsilon)$ , the expected value of the dipole measurement is

$$\langle s \rangle_\epsilon = \int_{-\infty}^{\infty} d\epsilon \, p(\epsilon) s(\epsilon), \quad (5.19)$$

where  $s(\epsilon)$  is the signal computed from Eq. 5.12 in the presence of an uncorrected error  $\epsilon$ . In order to assess the sensitivity of the aperture dipole statistic to these errors, we assume that position offsets and misalignment angles are both normally distributed, fully specified by their full-width-at-half-maximum (FWHM), and zero mean (i.e. no systematic spatial offsets or misalignments).

The sensitivity of the filters to decentering errors depends significantly on the resolution of the CMB experiment, as detailed in Appendix C. Positioning errors with  $\text{FWHM} = 0.2 R_{\text{vir}}$  ( $\sim 1$  arcmin), hardly suppress the signal for experiments that barely resolve galactic halos (such as *Planck*), and can yield a reduction of up to  $\sim 10\%$  for ACT. The aperture filter can be more sensitive than a matched filter, for high-resolution experiments, with a signal suppression of up to 40% for such a decentering on a CMB-S4 experiment.

The robustness of the aperture and matched filters to misalignments with respect to the true galaxies' atmospheres' spin angle does not depend strongly on the experimental configuration. Large errors in the spin angle with  $\text{FWHM} = 90$  deg result in a signal suppression of less than 25%.

### 5.6.3 Detection feasibility

The number of galaxies required for a  $3\sigma$  detection shown in Table 5.2 raises the question of how feasible is the measurement of a large number of galactic spins, and whether the rkSZ signal could be masked by other effects.

Table 5.3 shows the same information as Table 5.2, with number of galaxies converted to an equivalent survey sky coverage. To do so, we use the double Schechter local galaxy stellar mass function [252]:

$$n = \phi_1 \Gamma\left(\alpha_1 + 1, \frac{M_1}{M^\star}, \frac{M_2}{M^\star}\right) + \phi_2 \Gamma\left(\alpha_2 + 1, \frac{M_1}{M^\star}, \frac{M_2}{M^\star}\right), \quad (5.20)$$

	MaNGA-like	SAMI-like
	Hot, fast-rotating corona	
<i>Planck</i>	497   154	369   168
ACT	3   3	10   6
Simons	2   <1	7   1
CMB-S4	2   <1	6   <1
CMB-HD	<1   <1	<1   <1
	multi-phase, slow-rotating corona	
Planck	38,239   11,759	X   21,717
ACT	352   222	1,449   869
Simons	219   44	896   184
CMB-S4	194   3	789   12
CMB-HD	11   1	45   3

Table 5.3: Minimum footprint size of surveys required for a  $3\sigma$  detection of a rkSZ signal, in  $\text{deg}^2$ . This corresponds to the required number of galaxies, shown in Table 5.2, and converted to sky coverage using the stellar mass function described in § 5.6.3. An "X" indicates that the required sky area exceeds the full sky. The number on the left corresponds to the requirement using an aperture filter (measured at  $0.1 R_{\text{vir}}$ ) and the one on the right to the use of matched filtering.

where  $\Gamma$  is the incomplete gamma function,

$$\Gamma(x, a, b) = \int_a^b dt \exp(-t) t^{x-1}. \quad (5.21)$$

and the fitted values for the parameters are:  $\phi_1 = 4.26 \times 10^{-3} \text{Mpc}^{-3}$ ,  $\alpha_1 = -0.46$ ,  $\phi_2 = 0.58 \times 10^{-3} \text{Mpc}^{-3}$ ,  $\alpha_2 = -1.58$  and  $M^* = 10^{10.648} M_{\odot}$ .

For each of the survey bins shown in Figure 5.4, we multiply this galaxy number density by the corresponding comoving survey volume, taking into consideration the full sky solid angle ( $4\pi \text{ sr}$ ). The resulting number of available galaxies is  $5.76 \times 10^6$  for a MaNGA-like survey and  $6.58 \times 10^6$  for a SAMI-like survey. To have an idea of the depth required, the faintest galaxy in the MaNGA Primary sample has a magnitude of  $G=18.68$ . Upcoming CMB experiments should be capable of detecting the rkSZ signal with complete surveys of low-redshift galaxies covering a few thousands of  $\text{deg}^2$ .

In the absence of spectra, the orientation of the projected spin parameter could, in principle, still be estimated from photometry. From MaNGA data, the position angle of the single-component

Sersic fit in the  $r$ -band is highly correlated with the projected spin angle we estimated using the O II emission line. We find that the standard deviation of the difference between the photometric position angle and the O II derived angle is just  $9.0^\circ$ . However, the morphological direction from photometry alone leaves the sense of the rotation, which is crucial for attempts to measure the rkSZ signal, undetermined. On the other hand, the spectroscopic requirement to discern the sense of the projected spin, given its orientation, will be much less demanding than the ones reached in recent IFU surveys. Lower-resolution, faster surveys could be envisioned, maybe even narrow-band imaging, to detect the Doppler asymmetry on the two sides of a galaxy whose orientation is known.

Future HI surveys will provide additional information on gas kinematics for a large number of nearby galaxies, extending beyond their stellar component. For example, the WALLABY survey (Widefield ASKAP L-band Legacy All-sky Blind Survey,<sup>9</sup>) will cover  $\approx 31,000 \text{ deg}^2$ , up to a redshift of  $z < 0.26$ , and could marginally resolve  $\approx 540,000$  galaxies [253]. We expect that such marginally resolved observations could be used in isolation or combined with photometric surveys to break the spin-orientation uncertainty. Existing HI data has already been used to estimate galactic spins, showing that they are aligned with cosmic filaments [254].

Even with a number of galaxies' spin angle measurements sufficient to beat down the instrumental noise and the primary CMB anisotropies, alternative signals may mask that from the rkSZ effect. Two possibilities are the Birkinshaw-Gull effect (BG) and thermal dust emission.

The BG effect [255, 256, 257, 258] induces a temperature dipole, of a magnitude comparable to that of the rkSZ effect, with a decrease in CMB temperature following the galaxies' transverse proper motion (and a temperature increase opposite to the transverse proper motion). Regardless of the presence of systematic alignments between the directions of galaxies' spins and their proper motions —induced, for instance, by filaments [259, 260, 261, 254]—, we don't expect any systematic alignment in their sense. Thus, on spin-aligned stacks of large samples of galaxies, any BG-induced temperature dipole should average to zero.

---

<sup>9</sup>Widefield ASKAP L-band Legacy All-sky Blind Survey: <https://www.atnf.csiro.au/research/WALLABY/>

Another possible source of contamination is Doppler shifted thermal emission from dust co-rotating with the galaxy. There is evidence for the presence of dust in galactic halos [262] and it could co-rotate with the galaxies. Such a contaminant can be separated, in principle, using multi-frequency CMB measurements. In the worst-case scenario, dust emission could be used on its own as a tracer of galactic rotation in those regions within galactic halos where dust is kinematically coupled to gas. In the *Planck*’s high frequency stacks (see lower panels of Fig. 5.9), we see what appears to be an (unresolved) signal which would combine dust emission and tSZ effect from the field spiral galaxies (see a similar detection at galaxy cluster level in [263]). There is no clear indication of a Doppler-shifted dipole in these maps either.

The non-detection of a significant temperature dipole around field spirals using *Planck* data, puts a  $3\sigma$  upper limit on the average temperature dipole around field spirals of  $1.9\ \mu\text{K}$  (measured on a  $1\ R_{\text{vir}}$  aperture). In contrast, several studies have measured CMB temperature anisotropies aligned with the rotation axis of nearby spiral galaxies of a few tens of  $\mu\text{K}$  [225, 226, 227, 228]. This discrepancy may be due to the shape of the anisotropy (a highly concentrated temperature anisotropy can be significantly suppressed when unresolved by the detector’s beam), or may indicate that the nearby spirals with (large) measured temperature anisotropies are not representative of the MaNGA sample used for our analysis. The magnitude of the CMB temperature asymmetry measured around nearby galaxies is too large to be caused by the rkSZ, and we refer to [225, 226, 227, 228] for a brief enumeration of possible causes.

## 5.7 Conclusions

In this paper, we analyzed the feasibility of detecting the kSZ signal from the coherent rotation of the gaseous halos of galaxies. Such a detection would provide novel insight into the angular momentum distribution of gas, and more broadly, into galaxy formation. Our analysis is based on two models for galactic atmospheres: a fully ionized gaseous halo rotating at the DM halo’s circular velocity, which can be considered an upper bound, and a more realistic model consisting of a multi-phase gaseous halo rotating at a fraction of the DM halo’s circular velocity. As a proof-

of-concept, we stacked public *Planck* CMB data on the positions of  $\approx 2,000$  MaNGA galaxies after aligning these data with the galaxies' locations and spins, and scaling them to their halos' expected angular diameters. Our main findings can be summarized as follows:

- The number of galaxies required for a  $3\sigma$  detection with current CMB data is large, at best  $2.2 \times 10^4$ , beyond the largest set of galactic spin measurements currently available. The primary limitation is angular resolution in the CMB data, which only marginally resolves gaseous halos.
- Upcoming high-resolution, low-noise CMB experiments will significantly reduce the required number of galaxies. A galaxy survey measuring the spins of nearby galaxies covering  $\sim 10 \text{ deg}^2$  could be sufficient to rule out upper-bound models, and a few hundreds of  $\text{deg}^2$  should be sufficient to detect galaxies' rkSZ effect.
- The use of matched filtering can reduce the number of galaxies needed up to an order of magnitude for future CMB experiments. Such measurements can be sensitive to signal modeling, particularly in the presence of non-random CMB temperature anisotropies at the scales of galactic halos, such as those induced by their thermal SZ effect.
- A stacking analysis of *Planck* CMB data on the position of MaNGA galaxies rules out average non-random temperature dipoles aligned with the spin angles of field spirals down to  $1.9 \mu\text{K}$ . This may be inconsistent with asymmetries of up to  $\approx 100 \mu\text{K}$  measured in nearby spiral galaxies (e.g. M31).
- If Doppler shifted, anomalous thermal dust emission is responsible for the measured asymmetries in nearby spirals, as is claimed in a recent study [264], it could mask the rkSZ signal induced by gaseous galactic halos. This contaminant could be removed before searching for the kSZ signal. Alternatively, the emission from dust itself could be used to trace the kinematics of hot gas.

We conclude that the rotational kinematic Sunyaev-Zeldovich signal, imprinted on the cosmic microwave background by spinning hot gas in galactic halos, is a promising and novel probe of galaxy formation, and should be feasible to detect in future, high-resolution CMB surveys, combined with estimates for the spin orientations of  $\lesssim 10^4$  galaxies. A much larger number of galaxies would allow studies of the dependence of the angular momentum of the gas on galaxy properties.



## Chapter 6: Non-Gaussian information from weak lensing data via deep learning

### 6.1 Introduction

The analysis of multiple probes, including the Cosmic Microwave Background (CMB) and large scale structure (LSS), have yielded very precise estimates for the parameters that define the standard cosmological model,  $\Lambda$ -CDM [127, 124]. Early fluctuations in the CMB evolved through gravitational instability and formed the structures we observe in the late universe. The evolution of the matter distribution in the universe encodes rich cosmological information that can be mined to test the standard model and constrain the possible values for its defining parameters.

Over 80% of the matter in the universe is non-baryonic Dark Matter (DM), detectable through its gravitational effects. It contributes to gravitational lensing, distorting the shapes of background galaxies to an extent that is usually too small to be directly observed. Weak gravitational lensing (WL) can, nonetheless, be measured statistically through the correlation in the shapes of galaxies [161, 5]. The lensed galaxies' redshifts allow the reconstruction of the matter density field's evolution [265], making WL one of the most promising cosmological probes. Lensing measurements and their analysis in a cosmological context are an essential part of experiments such as CFHTLenS <sup>1</sup>, KiDS <sup>2</sup>, the Dark Energy Survey (DES <sup>3</sup>) or HSC <sup>4</sup>, and will be included in even wider ( $\approx 10\times$  larger) surveys (Large Synoptic Survey Telescope, LSST <sup>5</sup>, the Euclid mission <sup>6</sup> and the Wide Field Infrared Survey Telescope, WFIRST <sup>7</sup>).

---

<sup>1</sup><http://www.cfhtlens.org/>

<sup>2</sup><http://kids.strw.leidenuniv.nl/index.php>

<sup>3</sup><http://www.darkenergysurvey.org>

<sup>4</sup><http://hsc.mtk.nao.ac.jp/ssp/>

<sup>5</sup><http://www.lsst.org>

<sup>6</sup><http://sci.esa.int/euclid/>

<sup>7</sup><http://wfirst.gsfc.nasa.gov>

The large volume of upcoming datasets raises the question of how to extract all the cosmological information encoded in them. Non-linear gravitational collapse distorts the Gaussian character of the initial fluctuations. Thus, two-point statistics are insufficient to characterize weak lensing data and additional descriptors have been considered to extract additional information [178]. An alternative approach is to transform the data so that non-linearities become less important and it is easier to recover the information encoded in the transformed field (e.g. with the power spectrum). Logarithmic transformations have been proposed for the 3D matter density field [266] and the 2D convergence [267], as well as other local, Gaussianization transformations [268].

Overall, non-Gaussian statistics such as lensing peaks and moments involving gradients of the convergence field are promising, since they can improve parameter errors by a factor of 2-3 compared to using only second-order statistics [32, 41, 31, 49, 47, 50, 53, 52]. It is not clear where the extra information lies, or if all of it is accessible [269]. It has been investigated and partially understood only for lensing peaks, which derive some (but not all) information from underlying collapsed DM halos [159, 121, 160]. This halo-peak connection has inspired the development of approximate analytic models for peak counts [166, 170].

All these statistics compress the information in the original dataset, typically a map representing a noisy estimate of the projected matter density field, into a low-dimensional descriptor that can be used to infer the parameters that determine how the data was generated. An alternative approach is to use deep learning techniques, which have proven successful in a wide range of areas [62] to infer cosmological parameters directly from the uncompressed raw data.

Artificial neural networks (NNs) are pattern recognition algorithms, in which a series of processing nodes, capable of performing simple operations, are connected to each other in a network. The nodes of a NN are typically arranged in layers, with nodes in one layer connected to those in the next. Information is fed to the NN through the input layer, its outcome comes from the output layer, and all intermediate steps are called “hidden” layers. The strength of the connections is stored in a series of weights that can be adjusted to match a given output; this process is called “learning”. This quality allows the use of NNs for forecasting and inference. While we do not have

a full understanding on what drives NNs predictive power [270], they have been successfully used in Astronomy, from source detection and classification, to light curve analyses and even adaptive optics control (see reviews on NNs in astronomy in [271, 272]).

Convolutional Neural Networks (CNNs) are particularly well suited to work on datasets with spatial information, such as images, since the connection of their convolution layers' nodes to subsets of the data take advantage of the high correlation of nearby points imprinted by the locality of physical processes. Recently they have been used to infer cosmological parameters from the 3D matter density field [273], and have been found to outperform constraints estimated from its power spectrum. Weak lensing provides (in principle) an unbiased map of the projected matter distribution. One of the aims of this study is to assess if neural networks over-perform relative to the power spectrum when analyzing 2D WL data, as they do for the 3D matter field. Similar techniques have also been used to generate data with the same statistical properties as the output of physically-motivated simulations [111, 110].

A similar study has recently applied convolutional neural networks to weak lensing data for inference [113]. Our study shares the same motivation and reaches similar conclusions, but has some differences. While [113] focused on the ability of deep learning techniques to differentiate between models along a known  $\{\Omega_m, \sigma_8\}$  degeneracy,  $\Sigma_8 \equiv \sigma_8 (\Omega_m/2)^{0.6}$  [274], we focus on the parameters' constraints that can be inferred by extracting information through neural networks. To do so we trained our networks on a set of 96 cosmological models covering a large region of the parameter space (see Fig. 6.1 for the distribution of those models). Furthermore, we used different simulation techniques, the architecture of our network is different and we compared the neural network to a different set of observables (power spectrum and lensing peaks, instead of skewness and kurtosis). Finally, we restricted our analysis to noiseless data, leaving the analysis of the effect of shape noise for a follow-up study.

The paper is organized as follows, in § 6.2 we describe how we generated the data used to train and test the CNN, the architecture and training of the network, and the summary statistics used as benchmarks. In § 6.3 we compare the performance of the CNN to that of alternative summary

statistics, in terms of its predictive accuracy and the cosmological constraints that can be inferred. In § 6.4 we discuss the implications of our results and we summarize our conclusions in § 6.5.

## 6.2 Data

The goal of this paper is to assess the performance of CNNs predicting cosmological parameters from WL data. We do so comparing the network’s predictions with those that can be inferred from statistics measured on the maps, as well as the credible regions that can be inferred around the predicted parameters. In this section, we describe how the WL data used was generated, the design and training of the CNN, and describe the summary statistics measured on the WL data: the power spectrum and lensing peaks.

### 6.2.1 Mock convergence maps

Our initial data set consists of mock convergence ( $\kappa$ ) maps generated assuming 96 different values for the matter density  $\Omega_m$  and the scale of the initial perturbations normalized at the late Universe,  $\sigma_8$  (see Fig. 6.1). We adjusted the Dark Energy density to enforce flatness,  $\Omega_{\text{DE}} = 1.0 - \Omega_m$ , and kept the rest of the parameters constant: baryon density ( $\Omega_b = 0.046$ ), Hubble constant ( $h = 0.72$ ), scalar spectral index ( $n_s = 0.96$ ), effective number of relativistic degrees of freedom ( $n_{\text{eff}} = 3.04$ ) and neutrino masses ( $m_\nu = 0.0$ ).

We singled out the cosmology with  $\{\Omega_m = 0.260, \sigma_8 = 0.800\}$  as a fiducial to compute the covariance of the observables used to assess the performance of the CNN (see § 6.2.3). The density of the model sampling increases towards the fiducial and shows some correlation with the direction of the  $\Sigma_8$  degeneracy,  $\Sigma_8 = \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^{0.6}$ , as can be seen in Fig. 6.1. We refer the reader to [121], where this suite of simulations was also used, and our pipeline LENSTOOLS [143] for a detailed description of our sampling algorithm and simulation processing. We provide a summary here for convenience.

We evolved the matter density field using the  $N$ -body code GADGET2 [144]. For each cosmology we simulated a single volume from initial conditions computed with CAMB [194]. The

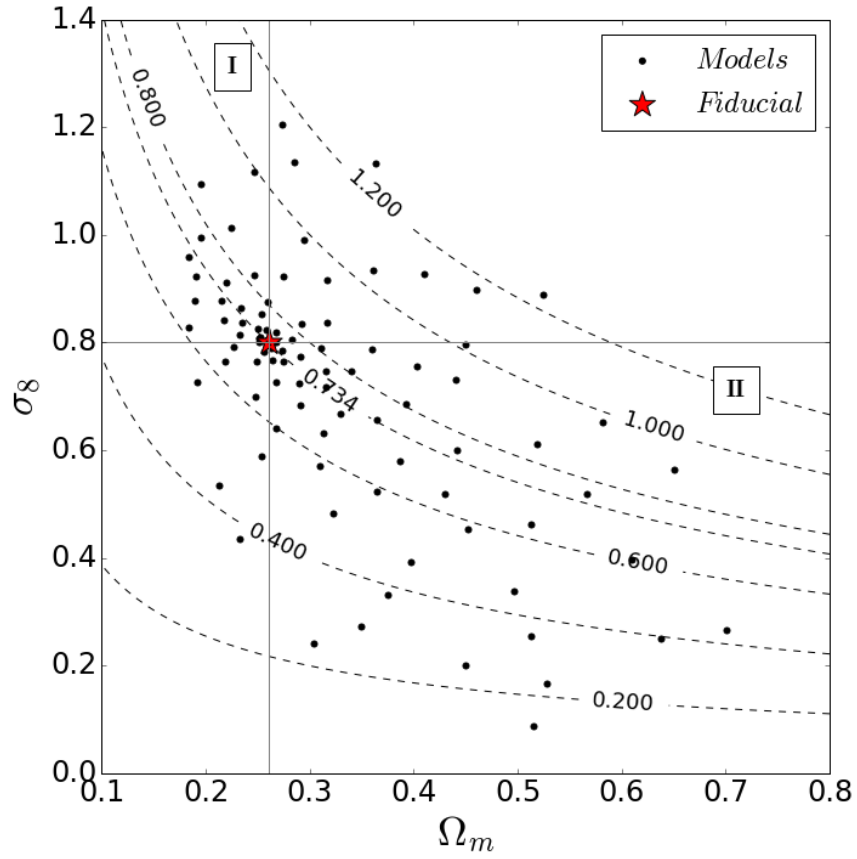


Figure 6.1: Location of the 96 cosmological models in our dataset on the  $\{\Omega_m, \sigma_8\}$  plane. The fiducial model,  $\{\Omega_m = 0.260, \sigma_8 = 0.800\}$ , is marked by a red star, and grey lines delimitate the quadrants defined by the fiducial parameters. The quadrants labeled **I** and **II** are discussed in § 6.4.3. The dashed curves show isolines for  $\Sigma_8 \equiv \sigma_8 \left(\frac{\Omega_m}{0.3}\right)^{0.6}$  for reference.

simulation boxes are cubes with a side-length of  $240 h^{-1}\text{Mpc}$ , large enough to cover the maps' field of view of  $3.5 \times 3.5 \text{ deg}^2$  to a redshift of  $z \approx 3.0$ . Each box is populated with  $512^3$  Dark Matter (DM) particles, yielding a mass resolution of  $\approx 10^{10} M_{\odot}$ .

We ray-traced the outputs of our simulations following the multiple lens plane algorithm [182]. It has been shown that while the Born approximation is sufficient for an accurate estimation of the power spectrum even in the largest planned future WL surveys, full ray-tracing is necessary to avoid biased estimations for the counts of lensing peaks and higher order statistics [142]. The value of  $\kappa$  for each of our maps' pixels is derived from the deflection experienced by a light ray as it crosses a series of lens planes stacked to form its past light-cone. For this study, we considered all the lensed galaxies located at a single fixed redshift of  $z = 1.0$ . Each resulting map has  $1024 \times 1024$  pixels, and was sliced in 16 smaller patches of  $256 \times 256$  pixels each to speed up the neural network's training (§ 6.2.2).

Each lens plane was generated from the snapshot corresponding to its redshift by cutting a  $80 h^{-1}\text{Mpc}$  slab along one of its axes, estimating the matter density on a  $4096 \times 4096$  grid, and solving the Poisson equation in 2D for the gravitational potential. By cutting different slabs, combining different planes at each redshifts, and randomly translating and rotating them, we ultimately generated 512 independent  $\kappa$  maps from a single simulation box for each cosmology. Through this recycling process, it is possible to generate up to  $\approx 10^4$  independent realizations of the convergence field from a single N-body simulation [143]. The resulting un-smoothed, noiseless convergence maps, is analogous to a 2D version of the dataset used in [273].

## 6.2.2 Neural network training and architecture

Neural networks consist of interconnected nodes (or neurons), arranged in layers. Each neuron transforms a linear combination of its inputs through an "activation" function,  $f(\mathbf{W}\mathbf{x})$ , where  $\mathbf{W}$  is a matrix of weights and  $\mathbf{x}$  a vector of inputs (in our case, the latter contains the values of the convergence in the pixelized 2D lensing map). The inputs can come from other neurons in the network, or from external data. The activation function is usually non-linear (e.g. a sigmoid

function). The weights used to linearly combine the inputs can be adjusted to minimize a loss function, in a process that is called “training” or “learning”. Some of the layers in the neural network used for this study convolve their input data with a kernel whose values are fitted during training. The resulting “convolutional neural network” takes advantage of the correlations between neighboring pixels and has been shown to yield good results when analyzing natural images.

Each “labeled example” the network is exposed to is a  $1024 \times 1024$  map coupled with the  $\{\Omega_m, \sigma_8\}$  “label” that corresponds to the cosmology used to generate that map. From each such example, we created 16 “labeled examples” by slicing the map into smaller,  $256 \times 256$  maps. And these are the maps used as input for the neural net. This operation reduced the number of nodes in the CNN and, consequently, its training time. We do not expect the performance of the network to be adversely affected, because of the limited constraining power of the modes that are small enough to be captured by the full maps but not their slices, i.e. spherical harmonic indices in the range  $\ell \in [100, 400]$  (see, e.g. ref [275], for a demonstration that most of the information is on smaller scales). The prediction for each  $1024 \times 1024$  map is the mean of the predictions for the 16,  $256 \times 256$  maps that were sliced from the original, bigger map. Our whole dataset amounts to 96 different cosmological models, each having 512,  $1024 \times 1024$  independent maps. We trained the neural networks using 70% of our data, and set aside the remaining 30% to test their performance.

The architecture of the CNN was inspired by that used in [273]. We sketch the architecture in Fig. 6.2 and summarize its elements in Table 6.1. The network is a combination of convolutions (transformed by a non-linear “activation” function) and pooling layers that reduce the spatial dimensions of the output, followed by fully connected layers in charge of the high-level logic. For the convolutional layers, we chose a  $3 \times 3$  kernel for speed. Each convolution layer applies more than 1 filter to its input in sub-layers. The weights (filter values) are the same for all the neurons within a sub-layer. This parameter-sharing reduces the number of weights to fit during training and is a reasonable choice given the data’s translational and rotational symmetries.

The first layer convolves any input map with 4 different filters and applies the activation function to the resulting 4 feature maps. Each filter is defined by 10 parameters (9 determine the

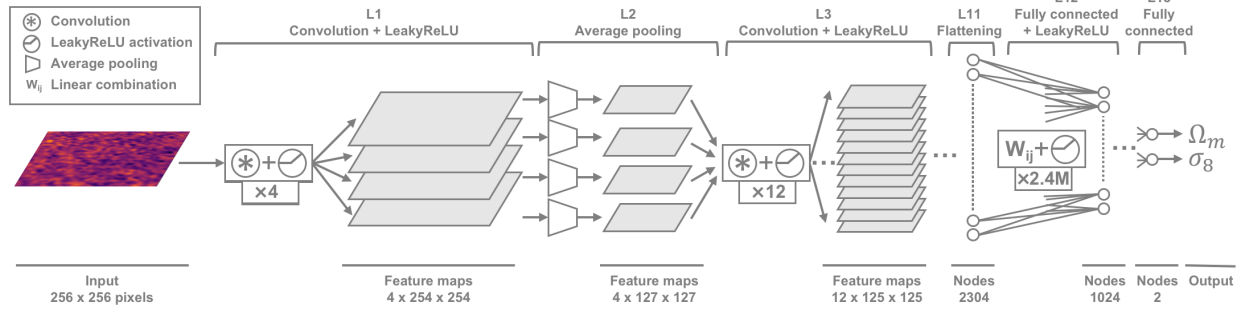


Figure 6.2: Schematic representation of the convolutional neural network (CNN) used in this study. The network consists of a series of convolutional and (average) pooling layers. Layers increase their “logical” dimension (depth), while reducing their “spatial” dimensions (width and height). Once the spatial dimension has been reduced to unity (flattening), a series of fully connected layers further reduces the number of nodes to two, the required number of outputs. The activation function for the neurons is a leaky rectified linear unit. For clarity, only a few layers are displayed.

convolution kernel plus an overall additive bias). In total, 40 weights need to be adjusted during training for the first layer. The second layer downsamples the feature maps from the first layer substituting  $2 \times 2$  consecutive pixels by their mean (“average pooling”). The third and fourth layers are convolutional layers, and each applies 12 different kernels to all incoming feature maps, including all depth levels from the previous layer. While the convolution is a linear operation, the application of the activation function breaks the linearity. The number of tunable weights grows with each layer as new kernels are added. Another average pooling layer (layer 5) is followed by two sets of convolution + average pooling (layers 6-9).

At each layer, we can consider the neurons arranged along 3 dimensions, 2 that follow the spatial dimensions of the feature maps fed into the layer (width and height) and another that grows with the number of filters used to process the layer’s input (depth). As information flows through the network, the spatial dimensions of the feature maps shrink and the depth of nodes processing those maps grows. The convolutional layers 6 and 8 do not apply an activation function to their output. Another average pooling (layer 10), followed by a flattening layer (layer 11) reduce the spatial dimensionality to unity, with a depth of 2304.

A series of fully connected layers (layers 12, 14 and 16) are followed by dropout layers (layers



Layer	Type	Sub-layers	Output dimension	Weights
1	Convolution + LeakyReLU	4	$4 \times 254 \times 254$	40
2	Average pooling	1	$4 \times 127 \times 127$	0
3	Convolution + LeakyReLU	12	$12 \times 125 \times 125$	444
4	Convolution + LeakyReLU	12	$12 \times 125 \times 125$	1308
5	Average pooling	1	$12 \times 61 \times 61$	0
6	Convolution	32	$32 \times 59 \times 59$	3488
7	Average pooling	1	$32 \times 29 \times 29$	0
8	Convolution	64	$64 \times 27 \times 27$	18496
9	Average pooling	1	$64 \times 13 \times 13$	0
10	Average pooling	1	$64 \times 6 \times 6$	0
11	Flattening	1	2304	0
12	Fully connected + LeakyReLU	1	1024	2360320
13	Dropout	1	1024	0
14	Fully connected + LeakyReLU	1	256	262400
15	Dropout	1	256	0
16	Fully connected + LeakyReLU	1	10	2570
17	Dropout	1	10	0
18	Fully connected	1	2	22
Total				2649088

Table 6.1: *Summary of the neural network’s architecture. Convolutional layers increase the depth of the network by applying different filters (sub-layers) to the same input. The number of neurons in a layer is determined by the dimension of its output. The number of weights for a convolutional layer is given by  $F_{out} (F_{in} \times 9 + 1)$ , where  $F_{out}$  is the number of feature maps that the layer outputs and  $F_{in}$  the number of feature maps the layer is fed with. A fully connected layer is defined by  $(N_{in} + 1) \times N_{out}$  weights, where  $N_{in}$  is the number of nodes in the previous layer and  $N_{out}$  the number of nodes in the fully connected layer.*

13, 15 and 17) that shrink the depth of the output. The final fully connected layer (layer 18) outputs the estimated values for  $\Omega_m$  and  $\sigma_8$ , which are compared with their true value through the loss function to adjust the weights in the network through back-propagation.

The total number of parameters to be fitted during training is  $\approx 2.6 \cdot 10^6$ , a large number but very small compared with the total number of pixels in the training data set ( $\approx 3.6 \times 10^{10}$ ).

The adopted activation function is the “leaky rectified linear unit” (LeakyReLU), with a leak parameter of 0.03, within the range suggested in [72]:

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0.03x & \text{if } x < 0 \end{cases} \quad (6.1)$$

This functions helps mitigate the “dying” ReLU problem, in which a neuron gets stuck in a region of zero gradient [71]. To prevent overfitting, we enforced regularization applying “dropout” at the fully connected layers: the output of any neuron was ignored with a 50% chance [276]. This process took part only during training, and the output from the nodes that were not dropped-out was doubled to compensate for the ignored neurons.

We used two loss functions to minimize during the training of our neural networks. The first one is the sum of the absolute error on  $\Omega_m$  and  $\sigma_8$ , computed over batches of 32 maps each, in which the data is split for each pass of the training examples:

$$\sum_{\text{map} \in \text{batch}} \left| \sigma_8^{\text{pred}} - \sigma_8^{\text{true}} \right| + \left| \Omega_m^{\text{pred}} - \Omega_m^{\text{true}} \right|. \quad (6.2)$$

This is a popular choice, and converges faster than the sum of the squares of errors because its gradient does not necessarily cancel near zero. Due to the heterogeneous sampling in parameter space of our simulated models, the network is exposed to fewer examples from cosmologies in sparsely sampled regions. This can induce a bias in the predictions. To assess the impact of the non-uniform sampling on parameter constraints, we also used a weighted loss function:

$$\sum_{\text{map} \in \text{batch}} W_{\text{cosmo}} \left( \left| \sigma_8^{\text{pred}} - \sigma_8^{\text{true}} \right| + \left| \Omega_m^{\text{pred}} - \Omega_m^{\text{true}} \right| \right), \quad (6.3)$$

where  $W_{\text{cosmo}}$  is a weight inversely proportional to the sampling density at the location of a cosmological model in parameter space. Errors in predictions for maps from cosmologies in sparsely sampled regions are more severely penalized than those for maps from densely sampled regions. We show in § 6.4.3 that such a weighted loss function reduces the bias in the predictions, at the cost of a longer network training, but has only a limited impact on the parameter constraints inferred from the predictions.

The algorithm used to minimize the loss function was an Adam optimizer [277] with a learning rate of  $10^{-4}$  and first and second moment exponential decay rates of 0.9 and 0.999, respectively.

We trained each network until the loss function converged, which took in most cases 5 epochs

(an epoch is a pass of all the training examples in the data set). The training maps were split in batches and randomly reshuffled after each epoch. The networks' weights were recomputed after each batch, minimizing the total loss over the 32 tiles. Each batch took 40 – 50 s on a NVIDIA K20 GPU with 5GB of on-board memory, at the NSF XSEDE facility <sup>8</sup>. To further reinforce the rotation-invariance of the dataset, all maps were rotated 90 deg with a 50% probability before feeding them to the network.

### 6.2.3 Alternative descriptors

In order to assess the performance of the CNN, we compared the accuracy of its predictions with that achieved through analysis of summary statistics. We used two observables, the power spectrum and lensing peak counts. Both compress the information available in a given WL map in a data vector of dimension small compared with the number of pixels in the original map.

The power spectrum is defined as the Fourier transform of the two-point correlation function of  $\kappa$  [5].

$$\langle \kappa(\boldsymbol{\ell}) \kappa^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta_D(\boldsymbol{\ell} - \boldsymbol{\ell}') P(\ell) \quad (6.4)$$

In the above expression  $\delta_D$  is the Dirac delta function and  $\boldsymbol{\ell}$  is the 2D angular wave vector. We measured the power spectrum on all 512 mock  $\kappa$  maps for each of the 96 cosmological models. We evaluated the power spectra on 20 bins, logarithmically spaced in the interval  $\ell \in [1 \times 10^2, 7.5 \times 10^4]$ . The minimum angular scale (maximum wavenumber  $\ell$ ) is set to prevent any loss of information at the pixel level. The finite resolution of our simulations results in deviations from theory at wavenumbers  $\ell > 5 \times 10^3$  with a significant loss of power for  $\ell \approx 10^4$ , as Fig. 6.3 shows for the fiducial cosmology.

The power spectrum is a widely used observable in cosmology, mainly because it fully characterizes Gaussian random fields and is a well-developed analytic tool. While the initial conditions for the matter perturbations are Gaussian (or nearly so), non-linear evolution introduces significant

---

<sup>8</sup><http://www.xsede.org>

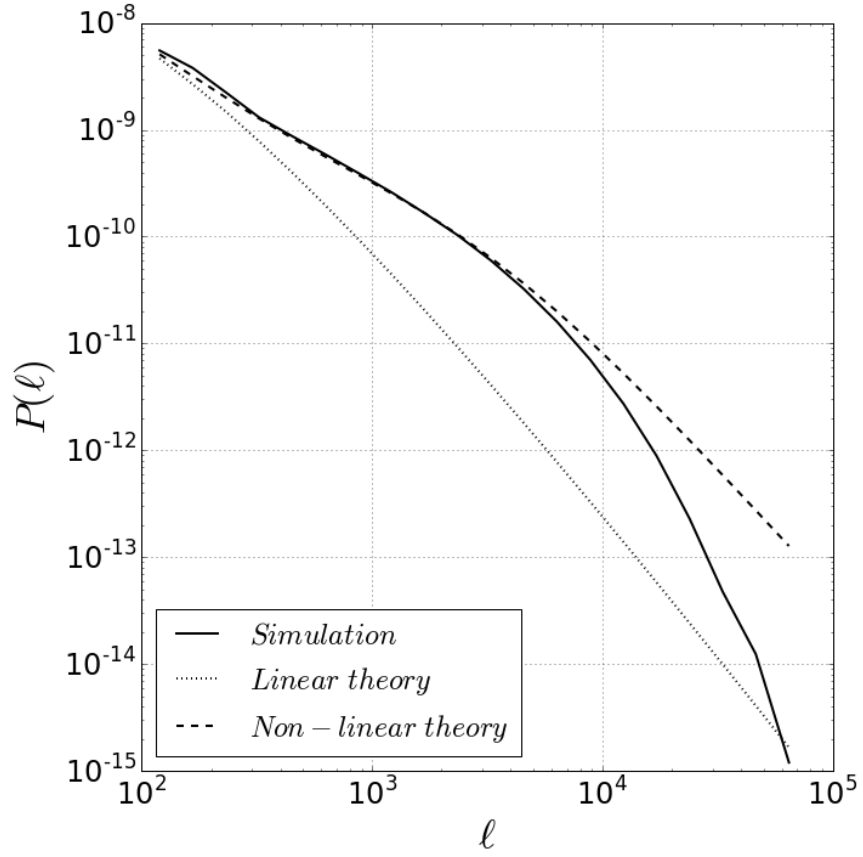


Figure 6.3: Comparison of the average convergence power spectrum for the fiducial  $\kappa$  maps with predictions from linear and non-linear theory. The theoretical curves were computed using NICA EA [278], with the revised Halofit parameters from [195] for the non-linear power spectrum.

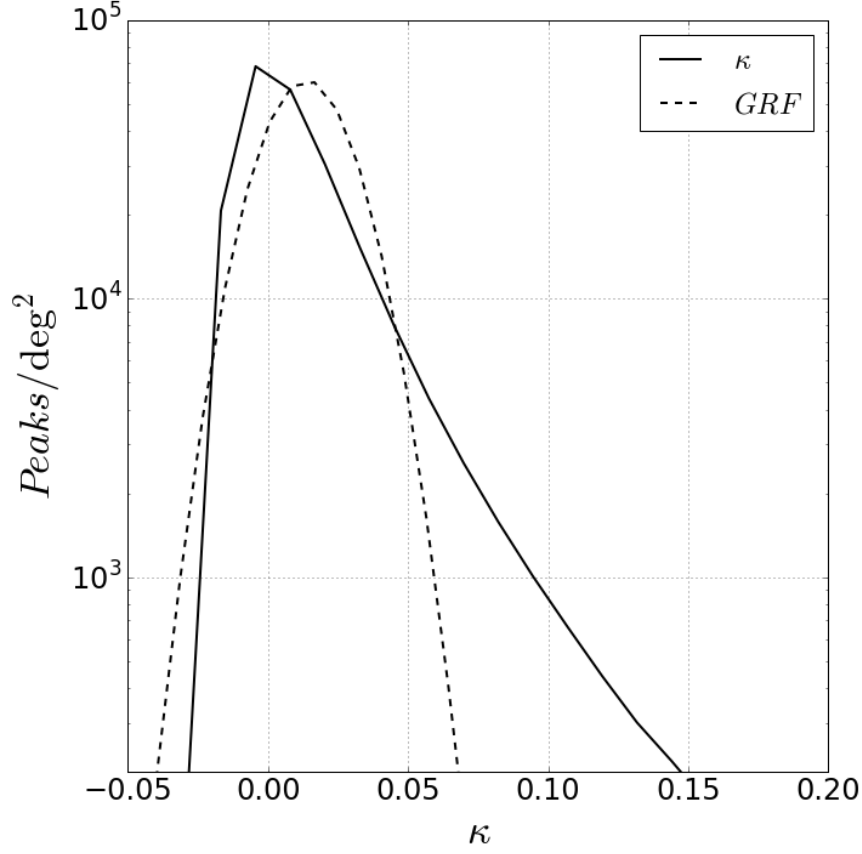


Figure 6.4: Comparison of peak counts derived from maps generated via our ray-tracing  $N$ -body simulations, to those derived from Gaussian random fields (GRFs) with the same power spectrum.

non-Gaussianities in the matter density field at late times.

Lensing peaks are local maxima in the  $\kappa$  field. In the absence of ellipticity noise, they probe high density regions, where non-linear effects become relevant. We chose the peaks' count as a function of their  $\kappa$  value as a second observable because they are sensitive to information not captured by the power spectrum. As an illustration, we compare in Fig. 6.4 the average peak counts measured on the 512 mock maps generated for the fiducial cosmology to those measured over Gaussian Random Fields (GRFs) that share their power spectra with the  $\kappa$  maps. That is, for each convergence map, we measured its power spectrum, built a GRF from it and measured the number of peaks in this new field. The distribution is clearly different, the peak histogram from convergence maps exhibiting a high  $\kappa$  tail resulting from the non-linear growth of structures.

Peak counts yield tighter constraints than the power spectrum [32, 31, 47, 53, 41, 49, 50, 52] and constitute a good benchmark for other methods which aim at extracting additional cosmological information. We counted the peaks in 20 bins, linearly spaced. We set the upper and lower limits of the bins to  $[\kappa_{min} = -2.0, \kappa_{max} = 12.0]$ , in units of the mean  $\kappa$  r.m.s. for the fiducial maps, to fully cover the range of peaks present in the data; this corresponds to  $\kappa_{min} \approx -0.03$ ,  $\kappa_{max} \approx 0.19$  and a bin width of  $\Delta\kappa \approx 0.01$ .

### 6.3 Results

We assessed the CNN’s performance in terms of the precision of their predictions for the cosmological parameters, and the constraints for those parameters for a given observation. The left and center panels of Fig. 6.5 display the predictions for  $\Omega_m$  and  $\sigma_8$  as a function of their “ground truth”, that is, the values that correspond to the cosmologies used to generate the data. The right panel shows the same comparison for the derived  $\Sigma_8 \equiv \sigma_8 (\Omega_m/0.3)^{0.6}$  along the degeneracy between both parameters. Each point corresponds to one of the  $\approx 150$  test maps available for each of the 96 cosmologies. For the neural network, the predicted  $\{\Omega_m, \sigma_8\}$  for a given map are the average values for the network’s output when fed the 16 tiles in which the map was sliced. For the power spectrum and peak counts, the predictions are the values that minimize  $\chi^2$  for that map. We estimated  $\chi^2$  for each of the 96 sampled cosmologies as:

$$\chi_{ij}^2 = (\mathbf{d}_i - \bar{\mathbf{d}}_j) \widehat{C_{fid}^{-1}} (\mathbf{d}_i - \bar{\mathbf{d}}_j) \quad (6.5)$$

where  $\mathbf{d}_i$  is the data vector measured on map  $i$  (binned power spectrum or peak counts),  $\bar{\mathbf{d}}_j$  is the mean of the same descriptor for the model  $j$  and  $\widehat{C_{fid}^{-1}}$  is the precision matrix for the data vector evaluated at the fiducial model. We used all 512 available maps per model to evaluate both the mean descriptor and the precision matrix, as in any realistic scenario in which a survey provides a mass map all the simulated data would be used for inference. We corrected for the bias in the

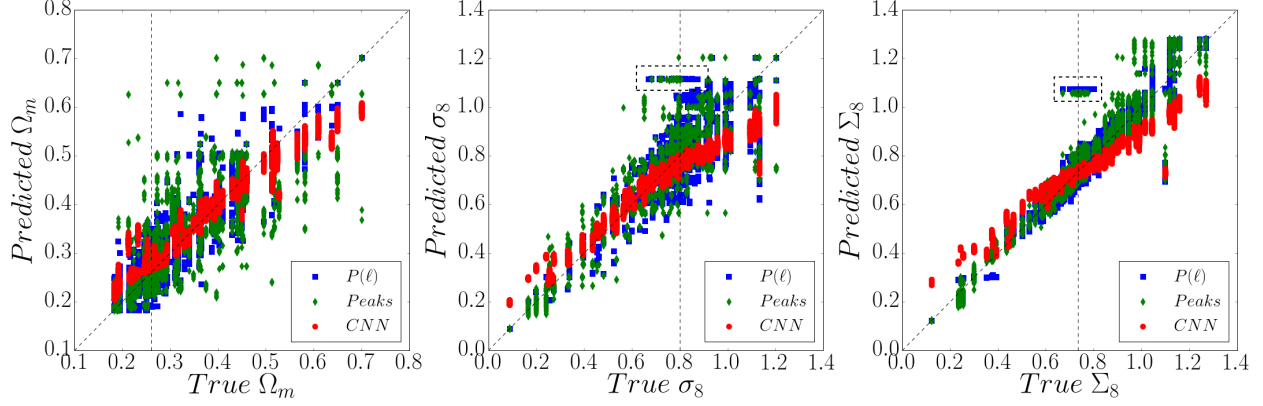


Figure 6.5: Predictions for  $\{\Omega_m, \sigma_8, \Sigma_8\}$  from un-smoothed ( $\approx 0.2$  arcmin/pixel) convergence maps, compared to their true values. Each point represents a map in the test data set. Predictions from the CNN are displayed in red, from the power spectrum in blue and from peak counts in green. Vertical dashed lines indicate the true values for the fiducial cosmology, and diagonal dashed lines the unbiased Prediction = Truth relationship. The dashed rectangles in the middle and right panels mark a small set of realizations of models near the fiducial cosmology; these contain anomalous structures leading to large biases (see text for discussion).

precision matrix following [151]:

$$\widehat{C_{fid}^{-1}} = \frac{N - d - 2}{N - 1} C_{fid}^{-1} \quad (6.6)$$

$N$  is the number of realizations used to estimate the covariance (512) and  $d$  is the dimension of the data vector (20, the number of bins).

The 96  $\chi^2_{ij}$  values were used to interpolate  $\chi^2(\Omega_m, \sigma_8)$  and find its minimum. We used a Clough-Tocher interpolator that builds a continuously differentiable piecewise cubic surface over a non-uniform grid [279, 280]. The minimum was found using the downhill simplex algorithm [281]. We verified that the results for the power spectrum and lensing peaks do not change when these observables are measured in a different number of bins (as long as they're more than  $\approx 10$ ) or a different interpolator is used to find the minimum of  $\chi^2(\Omega_m, \sigma_8)$ .

For all cosmologies, the neural network is significantly more precise than both the power spectrum and lensing peaks: the scatter in its predictions for a given model is smaller. On average, the standard deviation of the CNN's predictions is a factor of 4-7 lower than that of the statistical

	CNN	Power spectrum	Peak counts
	Noiseless, unsmoothed		
$\Omega_m$	5.1 (2.4)	21.7 (21.2)	35.6 (13.2)
$\sigma_8$	10.1 (7.9)	52.7 (84.1)	62.5 (63.8)
$\Sigma_8$	7.2 (4.7)	32.3 (73.2)	36.0 (59.2)

Table 6.2: *Standard deviation ( $\times 10^3$ ) of the predictions for the parameters  $\{\Omega_m, \sigma_8, \Sigma_8\}$ , averaged for all the cosmological models. In parenthesis, values for the fiducial model.*

descriptors, and up to  $\approx 16\times$  smaller for the fiducial (see Table 6.2). In terms of accuracy (i.e. how close the predictions are to the ground truth), the network shows some bias that may degrade the constraints that can be inferred from the network’s predictions.

We note the presence of a small set of maps from models close to the fiducial for which both the power spectrum and lensing peaks tend to over-predict  $\sigma_8$  and  $\Sigma_8$  as a result (the outliers on both panels correspond to the same maps). These maps form a clearly detached clump on the right-most panel of Fig. 6.5, where a dashed rectangle highlights their location. They represent  $\approx 4\%$  of the maps for  $\approx 28$  cosmologies not far from the fiducial model. We found through visual inspection that this over-prediction seems to be due to an anomalous number of structures projected in the field of view. Interestingly, the CNN seems to be immune to such chance projections and classifies these maps correctly. This suggests that the neural network extracts different information from the maps than the power spectrum or lensing peaks. Alternatively, these fluctuations may be the result from cosmic variance, and the neural network may be under-weighting those effects.

For a few cosmologies, parameter predictions from the CNN converged at different values from those of neighboring models. This is noticeable on the left-most panel of Fig. 6.5 where a few red points show a relative over-prediction in  $\Omega_m$  in the range  $\Omega_m \in (0.2, 0.4)$ . These outliers correspond to points in sparsely sampled areas near the boundaries of the explored parameter space. This highlights the importance of a well-sampled parameter space for the neural network to generalize accurately. In § B we analyze the effect of sampling on the predictions and credible contours inferred from the neural network. As these outliers lie far from the fiducial cosmology, they do not alter the parameter constraints presented in this study. Furthermore, they are identifiable in the



training data, and as such could be removed if needed. We did not remove any model from our data set even when it was evident from the training data that they could be outliers.

The relevant metric to compare the performance of the neural network relative to summary statistics is the probability distribution for the cosmological parameters given our data. This posterior distribution is related to the easier-to-compute probability of measuring a specific data vector given the cosmological parameters, or likelihood, by Bayes' theorem:

$$p(\mathbf{p}|\mathbf{d}, \mathcal{M}) = \frac{p(\mathbf{d}|\mathbf{p}, \mathcal{M}) p(\mathbf{p}, \mathcal{M})}{p(\mathbf{d}, \mathcal{M})}, \quad (6.7)$$

where  $\mathbf{p}$  is the set of cosmological parameters,  $\mathbf{d}$  a data vector and  $\mathcal{M}$  the underlying model, in our case DM-only simulations of  $\Lambda$ CDM cosmologies. For the CNN, we define our data vector as the predicted values for the cosmological parameters,  $(\Omega_m, \sigma_8)$ , and for the alternative statistics, the measured binned power spectra and peak histograms described in § 6.2.3.

The term that multiplies the likelihood, or prior  $p(\mathbf{p}, \mathcal{M})$ , and that on the denominator, or evidence, are the same when using the neural network or the statistical descriptors. The reason is we are using the same convergence maps from the same sampling of the parameter space. We can drop them as a normalization factor, as well as the explicit dependence on the underlying model used to generate the  $\kappa$  maps, and compare directly the likelihoods derived from the different methods. For the likelihoods, we assumed a Gaussian distribution:

$$p(\mathbf{d}|\mathbf{p}) \propto \exp \left[ -\frac{1}{2} (\mathbf{d} - \bar{\mathbf{d}}(\mathbf{p}))^T \widehat{C_{fid}^{-1}} (\mathbf{d} - \bar{\mathbf{d}}(\mathbf{p})) \right], \quad (6.8)$$

with a precision matrix  $\widehat{C_{fid}^{-1}}$  evaluated at the fiducial cosmology and as an expected value for the data,  $\bar{\mathbf{d}}(\mathbf{p})$ , the mean value measured from the simulations for the cosmology defined by  $\mathbf{p}$ .

Since we use the same covariance matrix for all cosmologies, we do not need to include the normalization pre-factor. For the power spectrum, we expect the Gaussian likelihood to be accurate. Our simulated maps cover a small field of view of  $3.5 \times 3.5 \text{ deg}^2$  on which the power spectrum can be measured only for relatively high  $\ell > 100$ . At those scales, many modes contribute to each

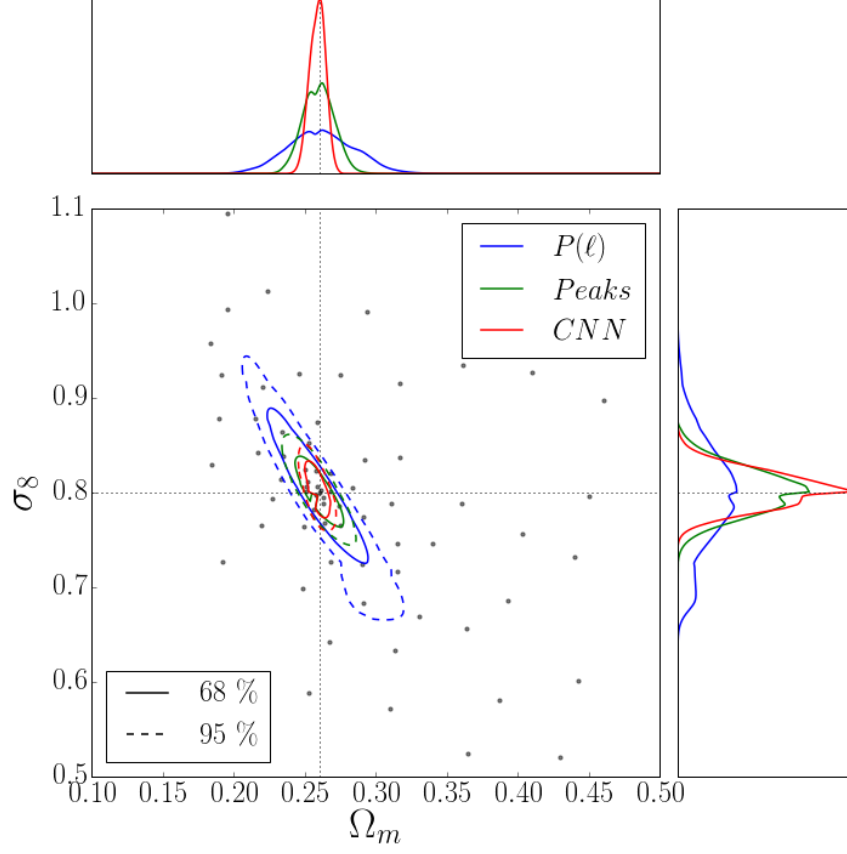


Figure 6.6: 68% and 95% credible contours for un-smoothed ( $\approx 0.2$  arcmin/pixel)  $\kappa$  maps, derived from the power spectrum (blue), lensing peak counts (green) and neural network predictions (red). The true values for the parameters,  $\{\Omega_m = 0.260, \sigma_8 = 0.800\}$  are indicated by black dotted lines. The upper and right panels show the distribution marginalized over the other parameter.

measurement of the power spectrum, and the central limit theorem shows that its probability distribution function should converge to a Gaussian [282]. For the lensing peaks and predictions from the neural network, we verified that the approximation remains valid (see § A). The alternative approach of estimating the probability density using a kernel density estimator (KDE) depends on the width of the kernel chosen, and the estimates for a large dimensional data vector such as our power spectra are noisy due to the relative limited amount of independent  $\kappa$  maps realizations.

To compute the likelihood, we used as data (observation) the average observable for the fiducial cosmology. For the power spectrum and lensing peaks, all 512 maps were used to estimate the means for each cosmology, and the covariance matrix for the fiducial. For the neural network, only

	CNN	Power spectrum	Peak counts
Area <sub>68</sub>	1	5.9	1.9
Area <sub>95</sub>	1	6.1	1.9

Table 6.3: *Area of the 68% and 95%  $\{\Omega_m, \sigma_8\}$  credible contours, relative to those obtained from the output of the neural network for un-smoothed, noiseless  $\kappa$  maps.*

the test maps were used ( $\approx 150$  per cosmology). We display the 68% and 95% credible contours for the likelihoods computed for the power spectrum, lensing peaks and neural network in the central panel of Fig. 6.6, and the marginalized distributions for  $\Omega_m$  and  $\sigma_8$  in the upper and right panels, respectively. At each point in parameter space, the expected data vector is interpolated linearly from the mean data vectors for the simulated cosmologies. Due to the choice of measurement (the predicted mean for the fiducial) all likelihoods peak at the true values for the fiducial cosmology. This is true also for the neural network. The smaller scatter in the CNN predictions translates into tighter parameter constraints, by a factor of  $\approx 2$  compared with lensing peaks and  $\approx 6$  compared with the power spectrum (see Table 6.3). The neural network seems capable of extracting more information from noiseless convergence maps than alternative methods such as the power spectrum or lensing peaks.

## 6.4 Discussion

### 6.4.1 Non-Gaussian information extracted by the neural network

The significantly tighter constraints obtained by the CNN, shown in Fig. 6.6, are encouraging and an indication that weak lensing maps encode more information than what is usually used for inference. Neural networks are capable of extracting some of it, at least more than the power spectrum and even more than some non-Gaussian statistics such as lensing peaks. Given the large number of parameters that need to be fitted during training, there is the risk that the gain in precision comes from a form of overfitting, in the general sense of making predictions based on irrelevant information [79].

For instance, a Gaussian random field, GRF, is fully determined by its power spectrum. As a

result, no other statistic or method used to extract information from it should out-perform the power spectrum. To test whether the neural network satisfies this limit, we built a collection of GRFs and used it as a new dataset to train and test the CNN’s architecture. We generated the GRFs by Fourier transforming random fields with a Gaussian distribution defined by the power spectra measured over the  $\kappa$  maps ray-traced from the outputs of cosmological N-body simulations. The new suite, which has a one-to-one correspondence with the original data, has no information encoded beyond the power spectrum.

The 68% and 95% credible contours from the power spectrum, lensing peaks and the newly trained CNN, as well as the marginalized distributions for  $\Omega_m$  and  $\sigma_8$  are displayed in Fig. 6.7, which is analogous to Fig. 6.6 but for GRFs instead of  $\kappa$  maps from N-body simulations.

As before, the likelihoods peak on the true parameter values for the fiducial and the contours appear centered around  $\{\Omega_m = 0.260, \sigma_8 = 0.800\}$ . The likelihood for the power spectrum is the same as the one computed for the convergence maps. The likelihoods for lensing peaks and the neural network are different, and their contours larger than those derived from the power spectrum. In particular, the contours from lensing peaks are 1.7 (1.4) $\times$  larger for the 68 (95)% contours, and those from the neural network 2.6 (2.0) $\times$  larger. This result is consistent with the absence of information beyond the power spectrum in the Gaussian Random Fields, and demonstrates that the small scatter in the parameters’ predictions from the neural network trained on convergence maps is not the result of a tendency to overfitting by its architecture or other spurious effects.

Comparing the  $\{\Omega_m, \sigma_8\}$  predictions with the ground truth for the test GRFs, as done in § 6.3 for the  $\kappa$  test maps, we see that there is both an increase in the scatter and the bias of the neural network’s predictions (see Fig. 6.8). Both effects drive the deterioration in the parameter constraints that can be inferred from those predictions. Furthermore, the neural network seems almost insensitive to  $\Omega_m$ , as the predictions for all the test GRFs scatter around the median  $\Omega_m$  for the 96 cosmologies. The CNN cannot easily distinguish between models with different  $\Omega_m$  and defaults to the value that minimizes the loss function. The use of an unweighted loss function in this analysis may also have some influence, but the same behavior is not seen on  $\sigma_8$ . The power spectrum and

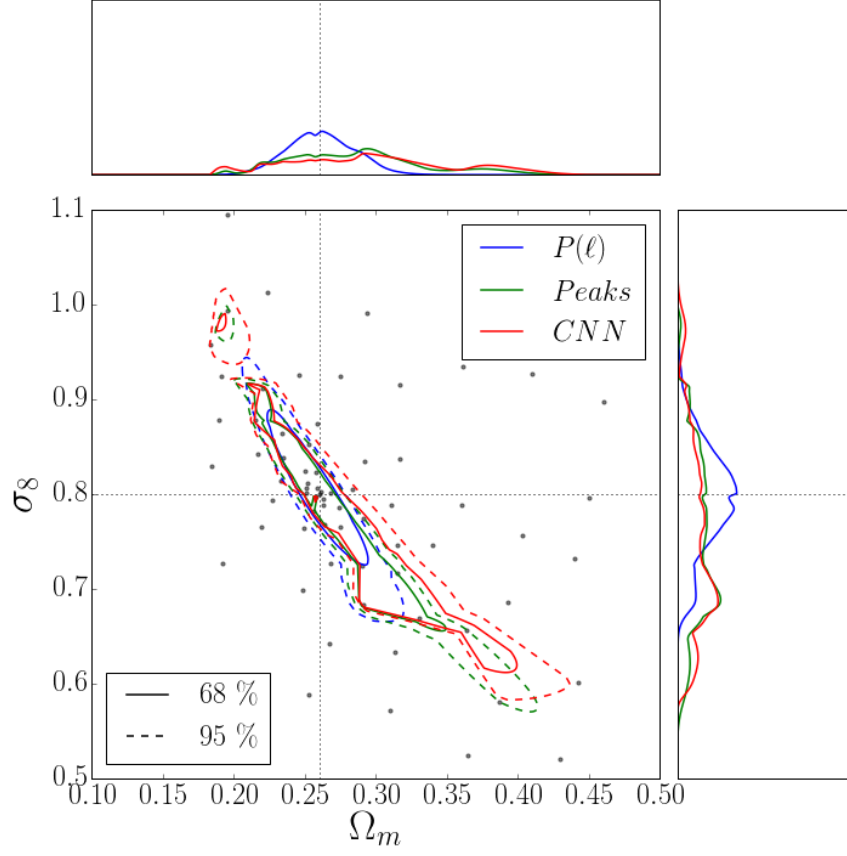


Figure 6.7: Same as Fig. 6.6, except using the Gaussian random fields, rather than the ray-tracing simulations. The network was trained with the un-weighted loss function (eq. 6.2).

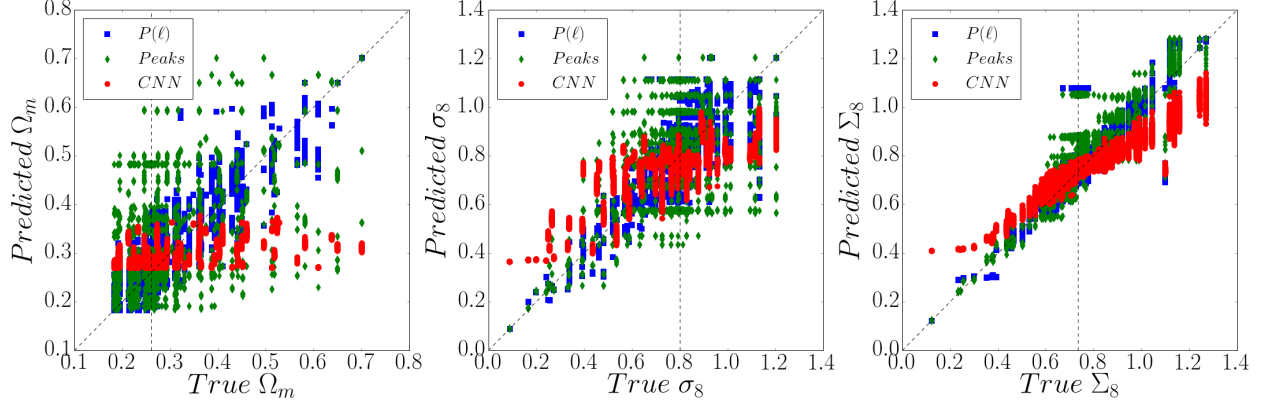


Figure 6.8: Same as Fig. 6.5, except using the Gaussian random fields, rather than the ray-tracing simulations. The network was trained with the un-weighted loss function (eq. 6.2).

lensing peaks are both sensitive to that parameter, indicating that they extract different information than the neural network.

#### 6.4.2 Effect of the smoothing scale on the results

The angular resolution of the mock convergence maps used for our analysis is  $\approx 0.2$  arcmin per pixel. This high resolution is interesting from an academic perspective, but at present it is of little practical interest. Accurate shear estimates require measuring the shape of many galaxies to estimate their correlations. For instance, the upcoming LSST survey will reach an effective number of galaxies of  $\approx 26 \text{ arcmin}^{-2}$ , after considering losses due to blending and masking [283]. This means that  $\approx 1$  arcmin is characteristic of the resolution achievable by future surveys. Furthermore, at small scales ( $\ell > 10^4$ ), baryonic physics alter the matter distribution and can bias WL observables relative to estimates from DM-only simulations [284, 285].

To assess whether the neural network still outperforms alternative observables on  $\approx 1$  arcmin resolution data, we trained a new network with the same architecture on the  $\kappa$  maps after smoothing them with a Gaussian kernel. The resulting constraints, for a smoothing scale of 1 arcmin, are displayed in Fig. 6.9.

The parameters' constraints degrade for all three methods. In principle, we would expect the

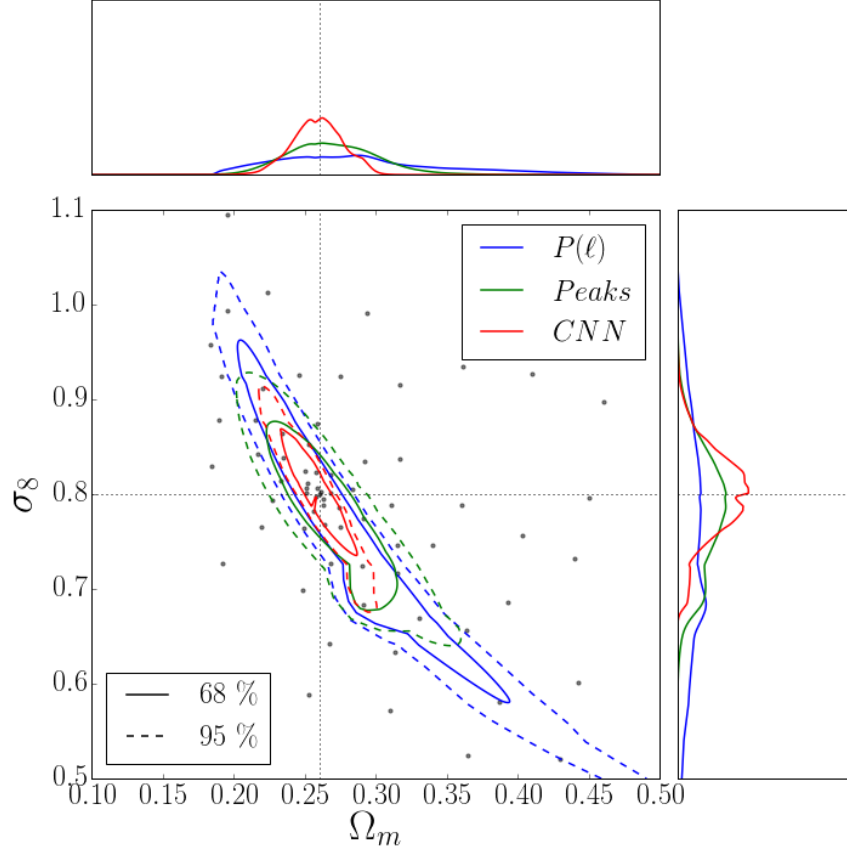


Figure 6.9: Same as Fig. 6.6, except smoothing the maps from the ray-tracing simulations with a Gaussian kernel of 1 arcmin of width. The network was trained with the un-weighted loss function (eq. 6.2).

Smoothing [arcmin]	Power spectrum		Peak counts	
	68 %	95%	68%	95%
-	5.9	6.1	1.9	1.9
0.2	7.0	5.9	1.8	1.6
0.3	7.7	7.9	2.3	2.7
0.4	6.5	6.4	1.9	2.2
0.5	7.1	6.5	2.5	2.5
0.6	6.5	5.7	2.5	2.4
0.7	6.4	5.2	2.8	2.4
0.8	4.7	4.1	2.5	2.3
0.9	5.2	4.4	3.0	2.8
1.0	5.6	4.8	3.6	3.3

Table 6.4: *Area of the 68% and 95%  $\{\Omega_m, \sigma_8\}$  credible contours, relative to those obtained from the output of the neural network, for different smoothing scales of  $\kappa$  maps. The first row corresponds to the un-smoothed data.*

non-Gaussian statistics’ performance to degrade relative to the power spectrum as small scale features are smoothed away from the  $\kappa$  maps. Up to 1 arcmin smoothing, the neural network keeps well its relative advantage to the power spectrum, yielding credible regions 5.6 (4.8) $\times$  smaller at the 68% (95%) level. Lensing peaks are more adversely affected than the CNN by smoothing, yielding contours that are only 1.6 (1.5) $\times$  smaller than the power spectrum. This would indicate that any additional information extracted by the neural network is not confined to very small angular scales.

The first attempt at training the neural network on smoothed data failed. To guarantee the convergence in the training process, we gradually smoothed the  $\kappa$  maps in a similar way as [113] added noise to theirs. We fed the network with maps of growing smoothing scale, starting with a kernel of 0.2 arcmin of bandwidth. Once the network reached convergence at a smoothing scale, the kernel’s bandwidth was increased by 0.05 arcmin and the network re-trained. In all cases the neural network kept its advantage (see Table 6.4). The ratio between the areas of the credible regions derived from the power spectrum and the neural network remained roughly constant, while the same ratio for the lensing peaks and neural network increased as the capability of peaks to extract information degraded faster with larger smoothing scales.



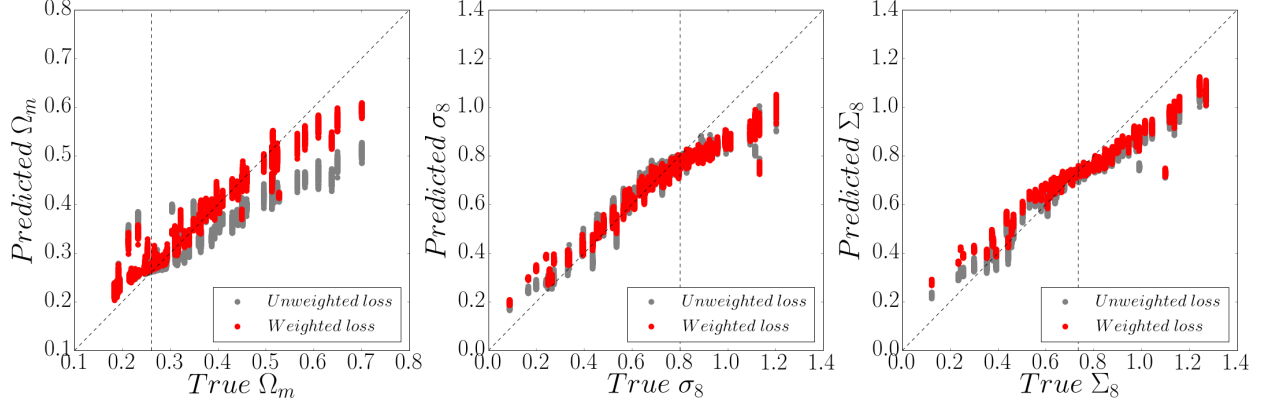


Figure 6.10: Predictions from the CNN for  $\{\Omega_m, \sigma_8, \Sigma_8\}$  from unsmoothed ( $\approx 0.2$  arcmin/pixel) convergence maps, compared to their true values. Each point represents a map in the test data set. Predictions using the unweighted loss function (eq. 6.2) are displayed in grey, and those using a weighted loss function (eq. 6.3), to account for the heterogeneous sampling of the parameter space, in red. Vertical dashed lines indicate the true values for the fiducial cosmology, and diagonal dashed lines the unbiased Prediction = Truth relationship.

#### 6.4.3 Bias in the CNN predictions

The parameter predictions from the neural network exhibit some bias (see Fig. 6.5). The bias is more severe when an unweighted loss function is used, as can be seen in Fig. 6.10. This can be due to the loss function being dominated by errors in the densely sampled regions of the parameter space.

Weighting the loss function according to the sampling density helps mitigate the bias. The effect is larger for the high- $\Omega_m$  region than for the high- $\sigma_8$  models. This can be due to the difference in sampling between both regions. The high- $\Omega_m$  region, corresponding to quadrant II in Fig. 6.1 has more models further from the fiducial and with large spacing between them than the high- $\sigma_8$  region (quadrant I).

The weights in the loss function were computed using a kernel density estimator (KDE) to estimate the sampling density in parameter space. The KDE bandwidth used was 1.0, a value that yielded a smooth estimate.

Biases in predictions from neural networks have been found in other works (e.g. [273]), so we cannot guarantee that the heterogeneous sampling of our data is the only source of the bias. Future

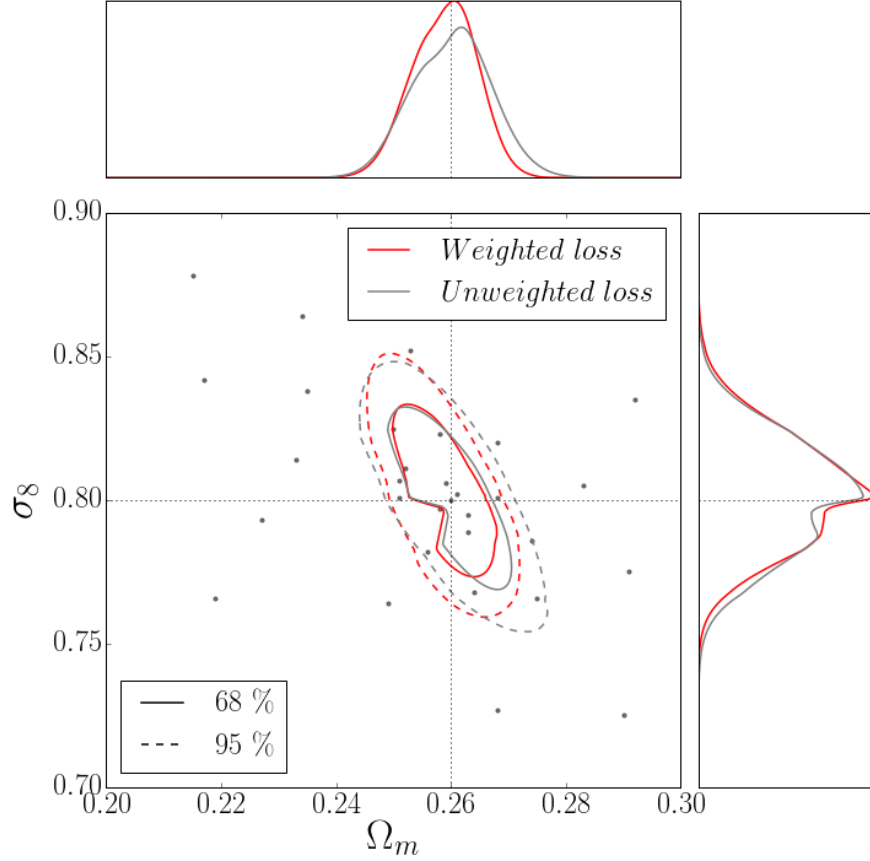


Figure 6.11: 68% and 95% credible contours for un-smoothed ( $\approx 0.2$  arcmin/pixel)  $\kappa$  maps, derived from two neural networks with the same architecture: in red the result from training with the weighed loss function (eq. 6.3) and in grey the result from training with the un-weighted loss function (eq. 6.2). True values are indicated by black, dotted lines. The upper and right panels show the marginal distribution for  $\Omega_m$  and  $\sigma_8$ , respectively.

work using a different dataset, uniformly sampled, will address this issue.

The parameter constraints for an observation near the fiducial model are not affected by the use of an unweighted loss function, as Fig. 6.11 illustrates. This is because the scatter of the predictions in densely populated areas does not increase significantly when the bias in the sparsely sampled areas is reduced with a modified loss function. We did not re-train our networks with a weighted loss function due to the additional computational cost, since the constraints from the network's predictions are essentially unchanged.

## 6.5 Conclusions

We trained a convolutional neural network on simulated, noiseless, weak lensing convergence maps. We demonstrated that neural networks can outperform methods based on traditional observables such as the power spectrum, or even statistics previously shown to extract non-Gaussian information, such as lensing peaks. On data smoothed at 1 arcmin scales, within reach of upcoming surveys, the neural network outperformed the power spectrum by a factor of  $\approx 5$  and the lensing peaks by a factor of  $\approx 4$  (using the area of the confidence contour in the  $\{\Omega_m, \sigma_8\}$  plane as a figure-of-merit).

We performed null tests to verify that the improvement in the parameter constraints reflect the network’s ability to extract additional information present in the WL data, and is not a numerical artifact (for instance, some form of overfitting). This sets a lower limit to the cosmological information encoded in noiseless lensing maps, whether this is also the case in more realistic, noisy data sets, remains an open question. The network’s constraints are limited by both the precision and bias of its predictions. Whether further improvements are reachable through a different network architecture, or a richer training data set, remains an open question and calls for further investigation.

Our results are consistent with previous findings in [273] for the 3D matter power spectrum and in [113] for the ability of neural networks to distinguish WL data generated from different cosmologies. Some of the questions that future work will address are:

- Effect of noise on predictive power. The presence of realistic levels of noise (e.g. shape noise) can pose challenges to neural network training [113]. It remains to be shown if the  $\approx 5\times$  improvement in parameter constraints compared with the power spectrum is achievable with noisy data.
- Propagation of systematics on constraints from neural networks. Before neural networks can be used to infer parameters from weak lensing data, we need to understand the effect of the systematics present in the data on the resulting parameter constraints.

- Scaling with survey area. Since neural networks' training time steeply increases with the map size, it is important to assess how the constraining power from their predictions scale with map size, and how the scaling compares with that for alternative methods such as the power spectrum.
- Network analysis. While the interpretation of feature maps from deep networks (see [113]) is not straightforward, it may provide valuable insights to design new summary statistics capable of extracting cosmological information from lensing observations.
- Improvements in the network's training and architecture. An extended exploration of training parameters (density of models in parameter space, number of independent examples per model, loss function, etc.) and architecture's features (convolutional kernel size, number of layers, etc.) will elucidate the effect of these choices in the resulting constraints.

## Chapter 7: Interpreting deep learning models for weak lensing

### 7.1 Introduction

The perturbed trajectories of photons propagating through an inhomogeneous universe result in (de)magnified and distorted images of background galaxies. The effect is in general very small, but can be detected through measurements over a large ensemble of galaxies. This weak gravitational lensing (WL) can be used to reconstruct the matter density field between us, observers, and the lensed background galaxies [4, 5]. The statistical properties of this field, and its evolution, can be used to test the standard  $\Lambda$  + cold dark matter ( $\Lambda$ CDM) cosmological model. WL measurements are particularly sensitive to two of the defining parameters of  $\Lambda$ CDM: the mean matter density of the universe,  $\Omega_m$ , and the amplitude of the initial perturbations that acted as seeds for the growth of structure, which can be measured in the local universe as  $\sigma_8$ . Recent estimates for those parameters using WL measurements hint at a possible tension with values inferred from observations of the cosmic microwave background [286, 287, 20, 288], strengthening the case for more precise measurements.

Upcoming galaxy surveys, such as the Vera Rubin Observatory (VRO) Legacy Survey of Space and Time (LSST [21]), the Euclid space mission [22], and the Wide Field Infrared Survey Telescope (WFIRST [24]), will provide those measurements in the near future. It is crucial to extract and use optimally all the cosmological information encoded in the measurements obtained by these experiments. Some of the data products will be mass maps of unprecedented angular resolution, with the projected matter density (convergence, or  $\kappa$ ) up to a certain redshift.

The standard method to estimate cosmological parameters from such maps, within a Bayesian framework, is to compress the information content of all the pixels into a single data vector, or summary statistic, for which a likelihood can be computed or sampled using simulations. These

summary statistics can be physically motivated (e.g. the use of the power spectrum, supported by the independent evolution of different Fourier modes in the linear regime of the growth of structure), or aim to describe the map morphology (e.g. the total length of isocontours, or second Minkowski functional of a 2D field). The choice of any particular statistic will, however, generically entail a loss of information.

An alternative approach is to bypass the design of summary statistics, and use a data-driven algorithm to map directly the pixels in a map onto the parameters of interest. Recent attempts of doing so with deep neural networks (DNNs) have shown that these algorithms can provide competitive parameter constraints. This has been demonstrated not only for simulated data [114, 115, 34], but also for WL survey data [117]. While DNNs are capable of learning complex non-linear relationships between data and the parameters that control the generative models behind the data, they are notoriously difficult to interpret. This is due the large number of fitted parameters (weights and biases) involved, and the depth of the many layers that comprise them in a DNN.

Previous studies have attempted to understand DNN models trained on WL data. The feature maps output by a model’s intermediate layers have been found difficult to interpret [113], and the same applies to intermediate convolution filters [118]. The analysis of the convolution filters on the first DNN layer has proven more fruitful, with at least one example [93] of filters that could be interpreted and used to design a new powerful summary statistic (the distribution of the radial profiles of local maxima, or peaks, in the maps). The limitation of these methods is that they do not take into account the impact of the identified filters on the model’s output, which can be complicated by non-linear interactions with other components of the DNN, and as a result cannot connect features in the input data space to the output of the networks.

The aim of the present study is to interpret a high-performing DNN trained on WL data [34] using state-of-the-art attribution methods from the field of image classification. These so-called saliency methods have been developed to understand the output of DNNs by providing an attribution or importance metric for each individual pixel of a given input datum. This is a fast-evolving field, and many such methods have been proposed [289, 88, 86, 290, 291, 83, 292, 82]. For an

in-depth review of the sub-field of explainable DNN models, we refer the reader to [80].

This paper is organized as follows. First, we describe the DNN we study and the data used to train it, in § 7.2. In § 7.3, we then assess the performance of the model relative to a combination of summary statistics typically used to analyze WL data to confirm that the model processes information not accessible to these statistics. In § 7.4 we evaluate a series of attribution methods for DNNs proposed in the literature, and select the more appropriate ones for our combination of model and data. We use the selected methods to study features in input space with the largest impact on the DNN’s output. Finally, we discuss our results and summarize our conclusions in § 7.5.

## 7.2 Model and data

The DNN model analyzed in this study is the one developed in [34]. It is an architecture that combines 2D convolutional layers (18) and average pooling layers (6) to map inputs consisting of simulated WL converge ( $\kappa$ ) maps into two parameters of interests. Each convolutional layer is followed by batch normalization, except for the last one. All activation functions are rectified linear units (ReLUs), and the network was trained using stochastic gradient descent and a mean absolute error loss function. We will often refer to this specific DNN, including its architecture and learned parameters, as simply “model”.

The data set used for training and evaluation of the model is a suite of simulated  $\kappa$  maps generated by ray-tracing simulations of the growth of non-linear structure. This data set has been used in past studies of deep learning applied to weak lensing [114, 93, 34], and we refer the reader to these references for a detailed explanation of how the different maps are generated. The full suite consists in synthetic convergence maps for 101 different cosmologies, each defined by a distinct pair of parameters  $\{\Omega_m, \sigma_8\}$ , corresponding to the mean matter density of the universe (in units of its critical density) and the amplitude of the initial perturbations normalized in the local universe. For each cosmology, 512 independent  $3.5 \times 3.5 \text{ deg}^2$  maps were generated.

In this first academic study, we focus on the interpretation of a model trained on noiseless maps. Each map has a size of  $512 \times 512$  pixels, with a linear angular size of 0.41 arcmin per pixel, and

was smoothed in real space using a Gaussian filter with  $\sigma = 1$  arcmin.

The data was split into a training and a test set, encompassing 70% and 30% of the maps, respectively. The network was trained so that its output predicts  $\{\Omega_m, \sigma_8\}$ , divided by the standard deviation of those parameters in the data set.

### 7.3 Network performance relative to alternative statistics

Past studies have shown that neural networks offer a discriminating power between cosmological models competitive with alternative statistics, such as the power spectrum [114, 115, 117], lensing peaks [114, 34], skewness or kurtosis [113] and a combination of the power spectrum, lensing peaks and Minkowski functionals (MFs [118]).

We begin by comparing the performance of the DNN model to a combination of the power spectrum, lensing peaks and Minkowski functionals, to assess if the DNN exploits information not accessible through traditional summary statistics of weak lensing fields.

The power spectrum is the Fourier transform of the two-point correlation function. It fully characterizes Gaussian random fields, such as the matter density field after recombination, and is a commonly-used statistic in cosmology. However, as gravitational collapse induces non-Gaussianities in the matter density field, additional statistics are needed to extract all the information encoded in lensing data sets.

Counts of local maxima as a function of their height, or lensing peaks, is a statistic that is simple to measure and has been shown to improve constraints derived from using the power spectrum alone by up to a factor of  $\approx$  two [32, 31]. Lensing peaks have also been successfully used to analyze survey data and to improve parameter constraints [47, 48, 53, 52].

Three different Minkowski functionals can be defined for two-dimensional fields [150, 206] by performing measurements over excursion sets defined by the points whose value exceeds a given threshold. The first one,  $V_0$ , measures the area of the excursion set, the second  $V_1$ , the total length of its boundary, and the third,  $V_2$ , its genus. They have also been shown to improve constraints derived from the power spectrum by a factor of up to  $\approx 2 - 3$  [40, 50, 207, 44].



We combined the power spectrum, lensing peaks and the three Minkowski functionals into a single data vector, and estimated the constraints on the parameters  $\{\Omega_m, \sigma_8\}$  assuming the Gaussian likelihood:

$$P(\mathbf{s}|\boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2} [\mathbf{s} - \bar{\mathbf{s}}(\boldsymbol{\theta})] \widehat{\mathbf{C}}^{-1} [\mathbf{s} - \bar{\mathbf{s}}(\boldsymbol{\theta})]^T \right\}, \quad (7.1)$$

where  $\mathbf{s}$  is the measured data vector,  $\bar{\mathbf{s}}(\boldsymbol{\theta})$  the expected value of the data vector in a cosmology defined by the parameter set  $\boldsymbol{\theta}$ , and  $\widehat{\mathbf{C}}^{-1}$  the estimated precision matrix, which we evaluate at the single (fiducial) cosmology defined by  $\Omega_m = 0.260$  and  $\sigma_8 = 0.800$  (there is no need to consider the pre-factor with the covariance determinant).

Each statistic was measured in 20 bins. For the power spectrum, we considered uniformly spaced bins in logarithmic space, with spherical harmonic index between  $\ell \in [100, 15000]$ , and for the lensing peaks and Minkowski functionals, uniformly spaced bins in linear space, between  $\kappa \in [-0.0235, 0.0704]$ , which corresponds to  $[-2, 5]$  in units of the measured r.m.s. of the  $\kappa$  field for the fiducial cosmology. The expected data vector at a point in parameter space not present in the simulation suite is computed using an emulator. The emulator is a 2D Clough-Tocher interpolator (as implemented in the Python SciPy library [293]), fitted on the mean values of the data vector measured on the test data set.

The data vector covariance was estimated from a new suite of simulations in the fiducial cosmology, consisting of 30,000 independent  $\kappa$  maps ray-traced through the outputs of 100 new, independent N-body realizations of the underlying matter density field. We verified that this number of maps suffices via cross-validation —the credible contours for  $\Omega_m$  and  $\sigma_8$  computed from a 3-fold subdivision of the data are indistinguishable from each other, indicating numerical convergence. The bias in the estimation of the inverse covariance is accounted for by applying the correction factor  $\frac{N-d-2}{N-1}$  (where  $N$  is the number of measurements and  $d$  is the dimension of the data vector [151]).

We treated the cosmological parameters  $(\Omega_m, \sigma_8)$  output by the DNN as just another summary statistic that can be added to the data vector, increasing its size from 100 to 102. This allowed us a

Credibility	Change in credible contour area [%]					
	PS	PC	V <sub>0</sub>	V <sub>1</sub>	V <sub>2</sub>	All
68%	-93	-88	-80	-68	-38	-20
95%	-93	-88	-80	-69	-38	-20

Table 7.1: Percentage change in the area of credible contours derived from different statistics when they are combined with the output from the DNN (see Fig. 7.1 for a graphical representation of those contours). The statistic used for the right-most column (labelled “All”) is a combination of all the statistics in the other columns: power spectrum (PS), lensing peaks (PC), and Minkowski functionals (V<sub>0</sub>, V<sub>1</sub>, V<sub>2</sub>). The change in area is defined as  $\Delta\text{Area} = 100 \left( \frac{\text{Area}_{w/\text{DNN}}}{\text{Area}_{w/o\text{DNN}}} - 1 \right)$ .

uniform treatment of the DNN output and the other summary statistics within the framework of a Gaussian likelihood (for a test of Gaussianity of a DNN output, see [114]). Neither the means nor the covariance estimates used any of the  $\kappa$  maps present in the network’s training data set.

In Fig. 7.1, we show the credible contours for  $\Omega_m$  and  $\sigma_8$  that can be derived using summary statistics, the DNN, or a combination of both. The constraints are displayed separately for each individual statistic, and a combination of the power spectrum, lensing peaks, and Minkowski functionals. The percentage change in the area of the credible contours achieved when the output of the DNN is incorporated, defined as  $\Delta\text{Area} = 100 \left( \frac{\text{Area}_{w/\text{DNN}}}{\text{Area}_{w/o\text{DNN}}} - 1 \right)$ , is reported in Table 7.1.

The improvement relative to the power spectrum and lensing peaks is very significant, as has been shown in past studies [114, 115, 34, 117]. The improvement relative to the Minkowski functionals is more modest, in particular compared with V<sub>2</sub>, which is by far the most constraining when measured on noiseless  $\kappa$  maps (the neural network can nevertheless still improve the constraints obtained from V<sub>2</sub> by 38%). When combining all the statistics together, the addition of the DNN predictions manages to reduce the area of the credible contours by 20%, implying that the DNN can extract information in the maps that is not accessible to the alternative statistics.

Since the DNN does not improve constraints from the combination of the summary statistics by a large factor, it is worth checking if the network is just learning the statistics. First, we looked at the Pearson’s correlation coefficient between the DNN’s output and the measured statistics, conditioned to the true value of the parameters  $\{\Omega_m, \sigma_8\}$ , that is, we computed the correlation for

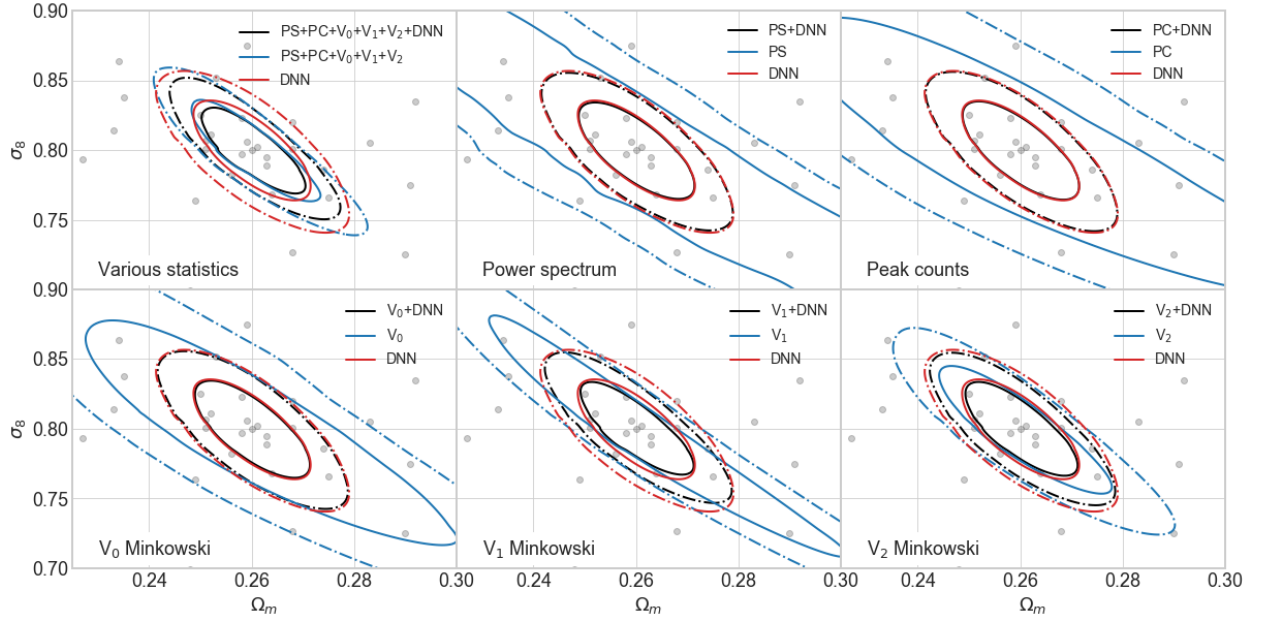


Figure 7.1: Credible contours derived for  $\Omega_m$  and  $\sigma_8$ . Each panel shows the comparison between the constraints derived from the DNN (in red) from an alternative statistic (in blue), and the combination of the DNN and the statistic (in black). Solid lines enclose 68% of the likelihood, and dot-dashed lines 95%. Upper row, from left to right: comparison between the DNN and a combination of statistics, the power spectrum (PS), and lensing peak counts (PC). Lower row, from left to right: comparison between the DNN and the three Minkowski functionals,  $V_0$ ,  $V_1$ , and  $V_2$ . The grey dots indicate the points in parameter space for which simulations were available.

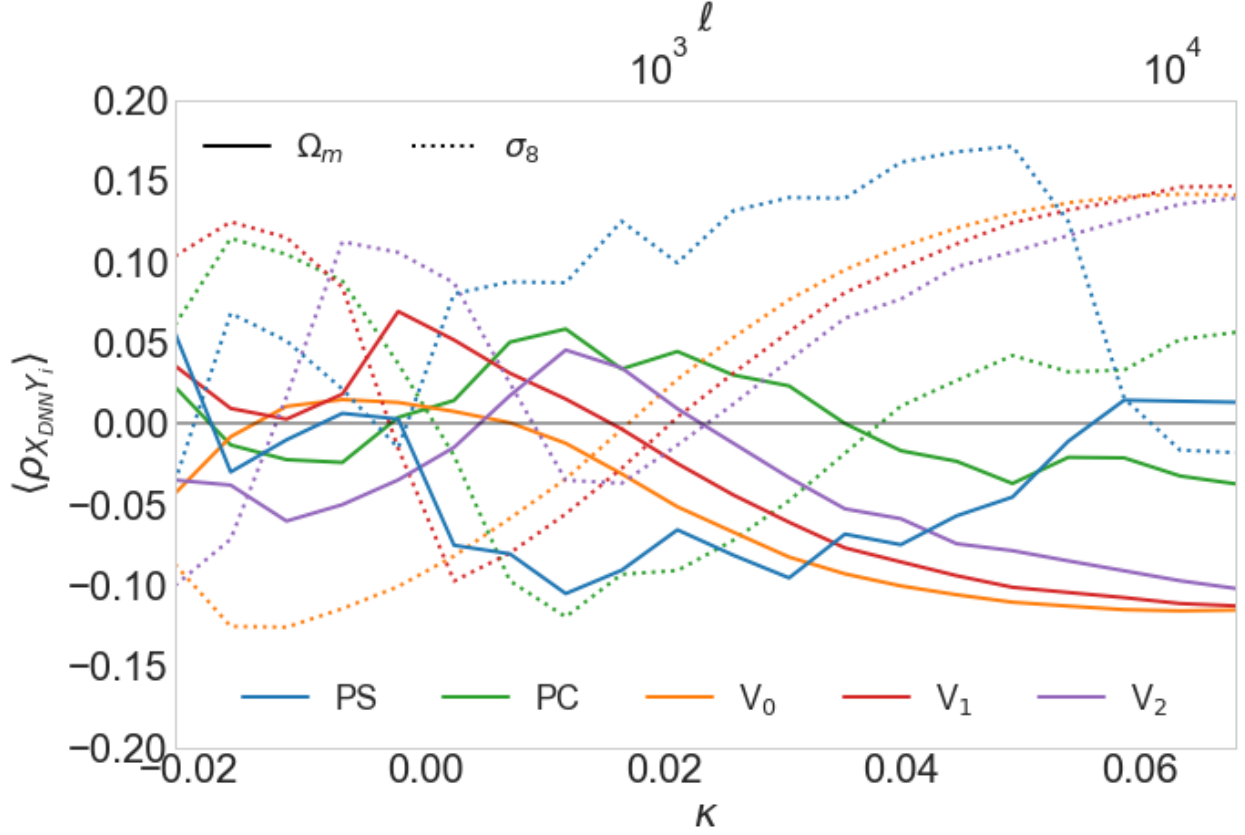


Figure 7.2: Pearson’s correlation coefficient (averaged over the 101 cosmologies) between the DNN predictions for  $\Omega_m$  (solid lines) and  $\sigma_8$  (dotted lines), and the measured statistics. For the power spectrum (PS, blue), the bins correspond to different multipoles ( $\ell$ ; see upper scale), and for the other statistics, values of  $\kappa$  (lower scale).

each cosmology in the test data set, and took the average of the values for the 101 cosmologies (the cosmology dependence of the correlations is weak).

Fig. 7.2 shows the average correlation coefficients as a function of  $\ell$  bin for the power spectrum (PS) and  $\kappa$  bin for the other statistics (computed from the test data set). None of the correlations is particularly high (all of the coefficients are below 0.2 in absolute value). The correlations for  $\Omega_m$  and  $\sigma_8$  tend to have opposite sign, indicating that the DNN learned the degeneracy between the two parameters. Besides, the qualitative change in the correlations as a function of binning follows expectations. For instance, higher  $\sigma_8$  is positively correlated with higher power spectrum, yielding a larger  $\sigma_\kappa^2$ , with results in a lower central peak and fatter tails for the lensing peak distribution.

It is also straightforward to show that the network’s output cannot be reproduced by a linear

combination of the summary statistics either. For each cosmology, we fit a linear combination of the summary statistics to the DNN output using the test data set. As we did with the correlations with the individual statistics, we averaged the coefficients for the 101 cosmologies (the cosmology dependence of the correlations is also weak). The resulting average correlations are also small compared to unity:

$$\begin{bmatrix} \rho(\Omega_m^{\text{DNN}}, \Omega_m^{\text{lin}}) & \rho(\Omega_m^{\text{DNN}}, \sigma_8^{\text{lin}}) \\ \rho(\sigma_8^{\text{DNN}}, \Omega_m^{\text{lin}}) & \rho(\sigma_8^{\text{DNN}}, \sigma_8^{\text{lin}}) \end{bmatrix} = \begin{bmatrix} 0.34 & -0.26 \\ -0.25 & 0.45 \end{bmatrix}$$

Given that (i) the DNN seems to access additional information and (ii) its outputs do not correlate highly with the summary statistics, or with their best-fit linear combination, we proceeded to look at the structure of the DNN to interpret its outputs.

#### 7.4 Interpreting DNNs with saliency methods

DNNs can be interpreted as non-linear mappings from an input space of dimension  $d$  (for this study,  $512 \times 512$ ) to a space of dimension  $n$  (for this study, two, the number of parameters of interest),  $\mathcal{M} : \mathbb{R}^d \rightarrow \mathbb{R}^n$ . Saliency methods map the input space into a space of the same dimension,  $\mathcal{S} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , so that the image of a given pixel,  $\mathcal{S}(x_i)$  is representative of the importance of that pixel for a given output neuron,  $\mathcal{M}(x_i)$ .

We analyzed several established methods that are well-defined for network architectures utilizing rectified linear units (ReLUs [71]), and do not require re-training of the model under study. These methods fall into two broad categories. The first category of saliency methods evaluate the effect of small perturbations of the input on the output. These methods rely on the the gradient of the DNN's output w.r.t. its input, which can be computed efficiently through a method called back-propagation—the iterative calculation of the gradient, layer-by-layer from the network's output to its input, avoiding redundant terms from the naive application of the chain rule [294, 70]. We selected two gradient-based methods whose interpretation for linear models is straightforward:

- **Gradient:** computes the gradient of the output neurons w.r.t. the values of the input pixels,  $\mathcal{S}(x) = \frac{\partial \mathcal{M}}{\partial x}$ . This measures the sensitivity of the output to the input, and for a linear model is equivalent to the regression coefficients.
- **Input×gradient:** computes the element-wise product of the input and the gradient of the output w.r.t. the input pixels,  $\mathcal{S}(x) = x \odot \frac{\partial \mathcal{M}}{\partial x}$ . For a linear model, it measures the contribution of the pixel to the output.

Other gradient-based methods exist, such as *Smoothgrad* [292] or *Integrated gradients* [295], but we did not study these methods due to their significantly higher computational cost. We inspected their effect on a small subset of input maps, and the results were qualitatively very similar to those of the *Gradient* and *Input×gradient* methods.

The second category of saliency methods tries to distribute the network’s output among the neurons of the second-to-last layer. The amount allocated to each neuron, interpreted as a relevance measure, is propagated iteratively through the network, back to the input space. We selected the following propagation-based methods:

- **Guided backpropagation:** masks out negative gradients and negative activations when back-propagating the gradient of the output w.r.t. the input [88].
- **Deconvnet:** uses a deconvolution network [296],  $\mathcal{M}^{-1}$ , built on top of the DNN architecture. To compute the saliency map corresponding to the input  $x$ , the feature maps  $\{f^i\}$  for each layer  $i$  in the model  $\mathcal{M}$ , are fed as inputs to the deconvolution network’s layers. At each stage of the propagation through  $\mathcal{M}^{-1}$ , intermediate representations are unpooled, rectified, and filtered, until pixel space is reached [87].
- **Deep Taylor decomposition:** distributes the relevance of neurons among its preceding layer by approximating the layer’s function with a first order Taylor expansion [290].
- **Layer-wise relevance propagation (LRP):** distributes the relevance of neurons among its preceding layer taking into consideration the weights that define the layer. We considered

two different rules that are common in the literature. The first one, LRP- $\epsilon$  uses as rule to propagate the relevance:  $R_i = \sum_j \frac{a_i w_{ij}}{\epsilon + \sum_i a_i w_{ij}} R_j$ , where  $a_i$  is the activation of neuron  $i$ ,  $w_{ij}$  the weight connecting neuron  $j$  to neuron  $i$ , the relevances  $R$  are the layer's output, and  $\epsilon$  absorbs weak or contradictory contributions to the activations. For ReLU-based networks [85],  $\epsilon = 0$  renders this method equivalent to *Input $\times$ gradient*. We chose  $\epsilon = 10^{-3}$ , for larger values resulted in saliency maps indistinguishable from random noise. The second rule, LRP- $\alpha\beta$ , propagates the relevance according to:  $R_i = \sum_j \left( \alpha \frac{(a_i w_{ij})^+}{\sum_i (a_i w_{ij})^+} - \beta \frac{(a_i w_{ij})^-}{\sum_i (a_i w_{ij})^-} \right) R_j$ , where  $()^+$  and  $()^-$  refer to positive and negative contributions. We used  $\alpha = 1$  and  $\beta = 1$ , a popular choice that renders this method equivalent to the *Excitation Backprop* method [79]. We also validated that our results do not change qualitatively when the parameters  $\alpha$  and  $\beta$  are modified slightly.

We applied the same method to all the layers in the DNN under study.

#### 7.4.1 Method comparison and selection

The column labeled “Trained model” in Fig. 7.3 shows the result of applying the different saliency methods to the same input map (shown in the left-most column labeled “Input”). As an illustration, we show the effect on a small  $0.68 \times 0.68 \text{ deg}^2$  path from a larger  $3.5 \times 3.5 \text{ deg}^2$   $\kappa$  map from the fiducial cosmology, but other regions and maps show the same characteristics. Also, for simplicity, all figures in this manuscript are built using the output of the DNN neuron that encodes its predictions for  $\Omega_m$ , but we have found identical conclusions when using the neuron corresponding to  $\sigma_8$ . To compute the saliency maps, we used the publicly available Python library *iNNvestigate* [297].

Visual inspection shows that the saliency maps from different methods can be qualitatively very different. Three of the propagation-based methods (*Guided backpropagation*, *Deep Taylor decomposition* and *LRP- $\alpha\beta$* ) show a clear correlation with structures in the input map, assigning high relevance to high  $\kappa$  regions, such as those around lensing peaks. The two gradient-based methods exhibit a more subtle correlation in which high- $\kappa$  regions have relatively low relevance, and high-

relevance peaks are instead associated with low- $\kappa$  regions around local minima of the input maps. The  $LRP-\epsilon$  map is very similar to the  $Input \times gradient$  map. We attribute this to the small  $\epsilon$  used (in the limit of  $\epsilon = 0$  the two methods are equivalent for our network architecture). Finally, the *Deconvnet* map exhibits some checkerboard artifacts, likely induced by the deconvolution scheme used [298], and little correlation with the input  $\kappa$  field.

Clearly, the different saliency methods provide very different answers to the basic question of "which input pixels are more relevant" to the DNNs output. It is therefore important to find a criterion to choose the method(s) most appropriate to interpret the model in the present context. Past work has shown that some saliency methods lack robustness [299, 300], and could be inappropriate for our combination of data and model. To assess the robustness of each method, we performed a model parameter randomization test, following the tests performed in [89]. For each method, we computed saliency maps not only on the trained DNN, but also on the models that result from randomizing the networks' parameters. We performed this randomization incrementally, starting with only the output layer, all the way to the first convolutional layer. Methods that yield saliency maps that are insensitive to these randomizations fail the test, as the structures in these saliency maps cannot then stem from features the DNN has learned during training.

As an illustration, the third and fourth columns of Fig. 7.3 (labeled "Last layer randomized" and "All layers randomized") show the saliency maps computed on the model after randomizing the weights of the output layer, and the weights of all the layers, respectively. Visually, the gradient-based methods (and  $LRP-\epsilon$ ) are very sensitive to the model's parameters, while propagation-based methods exhibit strong correlations between the saliency map computed on the trained and the random model.

To quantify the similarity between the saliency map computed from the model and from the model with all the layers randomized, we computed three measurements of association between both maps: the Pearson's  $r$ , Spearman rank-order correlation coefficient, and Kendall's  $\tau$ , reported in Table 7.4.1. For all three measurements, the null hypothesis is that there is no relationship (or correlation) between the two maps. The only methods for which the null hypothesis cannot be



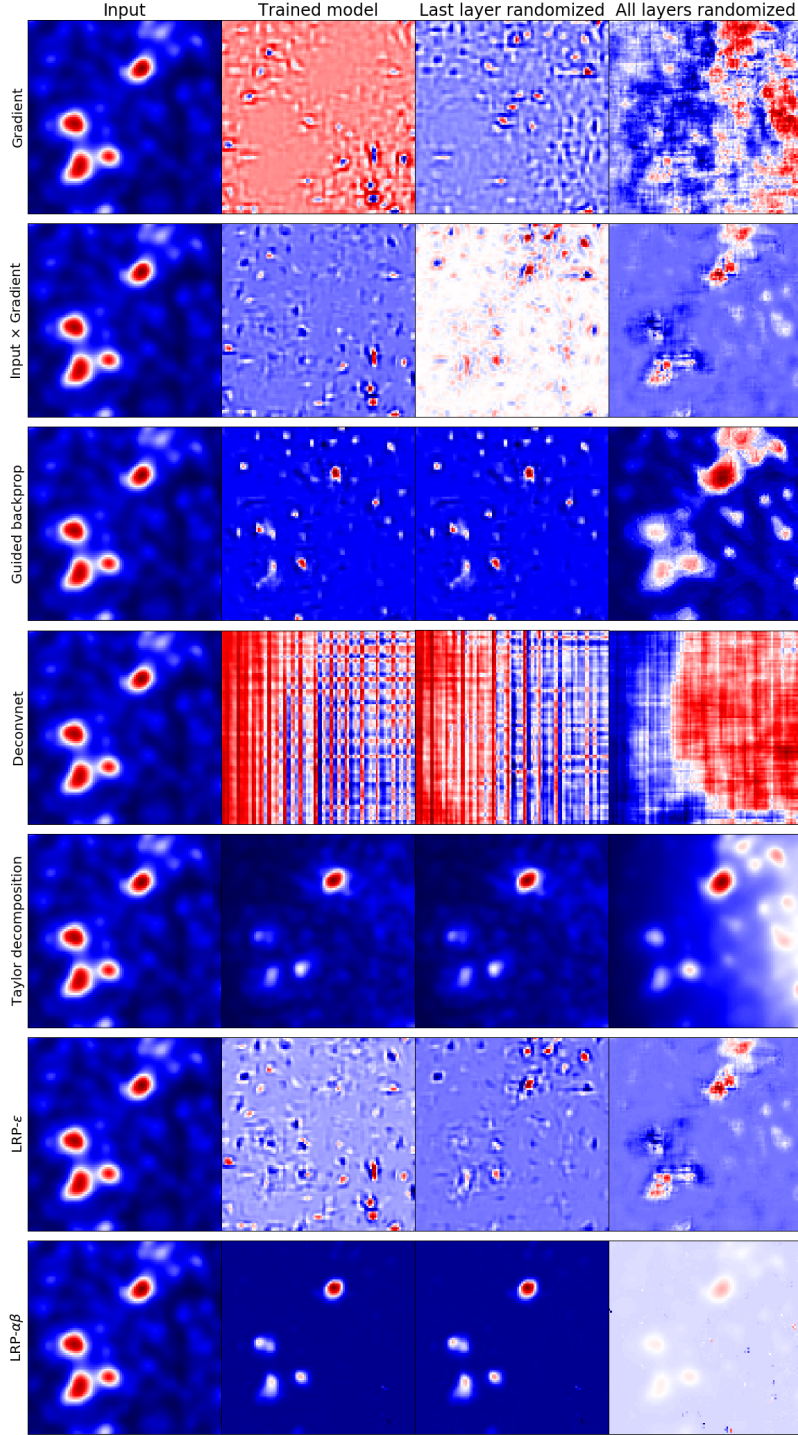


Figure 7.3: Examples of saliency maps for the output neuron of the DNN that encodes the parameter  $\Omega_m$ . The left-most column (“Input”) shows a small region ( $100 \times 100$  pixels, or  $0.68 \times 0.68 \text{ deg}^2$ ) of a  $3.5 \times 3.5 \text{ deg}^2$   $\kappa$  map from the fiducial cosmology. The second column (“Trained model”) shows the region of the saliency maps that corresponds to the region of the input map on the left. The third column (“Last layer randomized”) shows the same saliency map as the second column, computed on the fully trained model after randomizing the weights of the last (output) layer. The right-most column (“All layers randomized”) shows the same saliency map as columns 2-3, computed on a model where all the weights are randomized. Each row corresponds to a different saliency method. The scales for each image are omitted for clarity, since they do not influence the conclusions.

Saliency method	Correlation measurement					
	Pearson		Spearman		Kendall	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Gradient	-0.002	0.205	-0.001	0.527	-0.001	0.578
Input×gradient	-0.003	0.119	-0.000	0.944	0.000	0.785
Guided backpropagation	0.152	0.000	-0.025	0.000	-0.018	0.000
Deconvnet	-0.731	0.000	-0.675	0.000	-0.505	0.000
Deep Taylor decomposition	0.882	0.000	0.989	0.000	0.912	0.000
LRP- $\epsilon$	-0.011	0.000	-0.004	0.031	-0.003	0.024
LRT- $\alpha\beta$	0.463	0.000	0.173	0.000	0.115	0.000

Table 7.2: Correlation measurements between the saliency maps computed from the fitted DNN, and the same architecture with all the parameters randomized (see columns labelled “Trained model” and “All layers randomized” in Fig. 7.3). For all three tests (Pearson, Spearman, and Kendall), the null hypothesis is that there is no relationship between both saliency maps. P-values are two-sided.

rejected at high significance for all the three tests are the gradient-based methods. This is consistent with a past analysis of *Guided backprop* and *Deconvnet* [301]. Thus, in the rest of this paper, we will use the *Gradient* and the *Input×gradient* saliency maps to interpret the DNN.

#### 7.4.2 Mapping attributions back to physical space

We used the *Gradient* and *Input×gradient* methods to analyze the distribution of the relevance for the DNN output, as a function of  $\kappa$ . For each saliency map in our data set, we measured sum of the square of the pixel values (to avoid cancellations by gradients or inputs of different sign) for the pixels within a given range of  $\kappa$  values in the input map. We selected 20 linear bins with  $\kappa \in [-0.029, 0.0587]$ , or  $\kappa \in [-2.5, 5.0]$  in units of the  $\kappa$  r.m.s. for the fiducial cosmology  $\{\Omega_m = 0.260, \sigma_8 = 0.800\}$ . This measurement gives an estimate of the distribution of relevance in input space as a function of  $\kappa$ . We also measured the average relevance per pixel in each  $\kappa$  bin. These two measurements, for each of the 101 cosmologies in the data set, are displayed in Fig. 7.4. The color of each line corresponds to the value of  $S_8 = \sigma_8 \left( \frac{\Omega_m}{0.3} \right)^{0.6}$  in each cosmology; this is approximately the best-measured combination, orthogonal to the direction of the degeneracy between the parameters  $\Omega_m$  and  $\sigma_8$ .

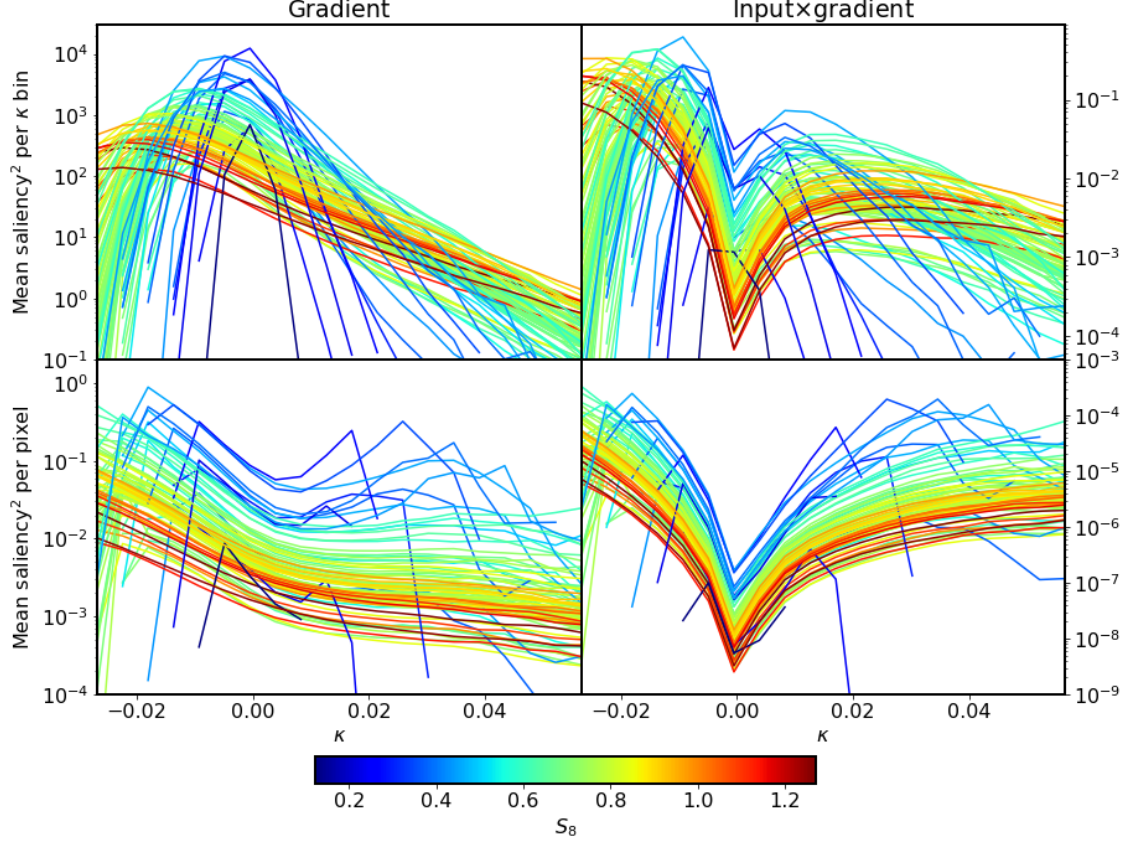


Figure 7.4: **Upper panels:** sum of the square of the pixel values in saliency maps as a function of  $\kappa$  in the corresponding input maps. Each line is the test maps' average for one of the 101 cosmologies. **Lower panels:** same as upper panels, divided by the number of pixels in each  $\kappa$  bin, giving the mean saliency<sup>2</sup> per pixel as a function of  $\kappa$ . Left panels correspond to saliency maps computed using the *Gradient* method, and right panels to saliency maps computed using the *Input×gradient* method. Each line is colored based on the value of  $S_8 = \sigma_8 \left( \frac{\Omega_m}{0.3} \right)^{0.6}$ .

The most relevant pixels, according to both gradient-based saliency methods, are those with extreme  $\kappa$  values. Those at the negative tail of the  $\kappa$  distribution are more relevant than those at the positive tail (see panels in the lower row of Fig. 7.4). These pixels are rare, and the most relevant  $\kappa$  regions are shifted towards the center of the distribution (see panels in the upper row of Fig. 7.4). Most of the relevance ends up being concentrated in regions with negative  $\kappa$ . These regions account for 89% of the sum of the squared pixel values in the *Gradient* saliency maps, and 86% in the *Input×gradient* saliency maps. We note that the drop in relevance around  $\kappa = 0$  in the results from the *Input×gradient* method is an artifact due to the zero input.

## 7.5 Discussion and conclusions

In this study, we analyzed in detail a deep learning model that has been shown in previous work to learn cosmological parameters from simulated noiseless WL maps smoothed at an angular scale of 1 arcmin. Our aim was to understand which features in the simulated WL maps are used by the model to derive its predictions. First, we compared its performance with a suite of statistics commonly used in the WL community, individually and in combination, and evaluated the correlations between the DNN output, those statistics, and their linear combination that best fits the DNN output. Second, we borrowed a series of saliency methods from the field of image recognition and applied them to the DNN trained on simulated WL maps. We tested each method, and selected those that passed a null test of robustness, showing that they are sensitive to the learned weights of the DNN model, and are not directly derivable from the input maps. Finally, we used these methods to identify which pixels in simulated WL maps does the DNN use preferentially to discriminate between cosmological parameters. Our key findings are the following:

- We generated a new suite of 100 simulations to measure accurately the covariance of the combination of five WL statistics (the power spectrum, lensing peaks, and three Minkowski functionals). Found that for noiseless, single-redshift simulated maps at 1 arcmin resolution, the third Minkowski functional,  $V_2$  is by far the most sensitive to cosmology.
- The DNN can extract information not accessible through a combination of the power spectrum, lensing peaks, and Minkowski functionals. The addition of the DNN to those statistics reduces the credible region on the cosmological parameters of interest by 20%.
- The DNN predictions are not highly correlated with the alternative statistics considered, nor can be reproduced using a linear combination of them.
- Saliency methods based on the back-propagation of the DNN output to input space were found to fail a simple robustness test: they are not sensitive to the values of the parameters that define the DNN. As a result, while (some) can provide attractive explanations in the

form of attribution maps that highlight structures present in the input data, they do not tell which features are learned by the model.

- Gradient-based methods are sensitive to the parameters learned by the model, and as a result they are safe to use to interpret which features the DNN learns from the data. Another advantage of these methods is that their interpretation, for linear models, is straightforward: they correspond to regression coefficients or measure the contribution of each pixel to the output.
- Gradient-based methods show that the most relevant pixels for the DNN are those with extreme values, at the tails of the  $\kappa$  distribution. Negative  $\kappa$  pixels are more relevant than positive  $\kappa$  pixels, and when the number of pixels is taken into account, most of the relevance for the model output lies in regions with  $\kappa < 0$  (86-89%).

The last conclusion is potentially our most astrophysically relevant result. Our analysis is based on simplified, single-redshift, noise-free simulated lensing maps, and neglect any systematic errors, and must be followed up with work addressing these aspects. If indeed negative  $\kappa$  regions provide the bulk of the cosmological sensitivity, this would have implications for the analysis of large future WL datasets. One encouraging aspect of this finding is that large voids, accounting for these demagnified / underdense regions, have been shown to be less affected by baryonic physics, which are hard to capture accurately in simulations of growth of structure [38, 39]. On the other hand, these regions have been shown to be sensitive to neutrino physics and modified gravity theories [35, 36, 37].

## Chapter 8: Conclusions and future work

Weak lensing as a cosmological probe is on the verge of a qualitative jump as Stage-IV experiments come online. In order to realize its full potential, it is critical to extract all the information encoded in the billions of galaxy shapes that will be measured. Doing so will require efficient statistics that can extract non-Gaussian information encoded on small,  $\sim \text{arcmin}$  scales, and forward models capable of predicting those statistics accurately for a wide range of cosmological models. In this thesis work we set out to address a few of the many aspects of this problem that need to be understood to facilitate this. As a conclusion, we summarize the key results presented in the previous chapters, and discuss possible future extensions of this work.

### 8.1 Summary of results

In Chapter 2, we showed how the sensitivity of non-Gaussian observables to geometry and growth have opposing signs, somewhat reducing their overall sensitivity to  $\Omega_m$  (for which growth dominates) and  $w$  (for which geometry dominates). Despite that partial cancellation, both effects help alleviate the degeneracy of WL measurements to these two parameters. The reduction in sensitivity can reach a factor of  $\approx 3$  for  $w$  and just  $\approx 1.5$  for  $\Omega_m$ , and the bispectrum is more affected than lensing peaks or Minkowski functionals.

In Chapter 3 we assessed the performance of a proposed fast emulator for lensing peak counts based on the halo model (CAMELUS), and found that it derives most of its sensitivity to cosmology from higher significance peaks than full simulations of  $\kappa$  maps. It over-estimates counts on the tails of the distribution, and under-estimates the covariance matrix by  $\approx 30\%$ , yielding optimistic parameter constraints. We suggested changes to take into account the spatial correlation in the position of halos for an upgraded version of the emulator.

As the numerical simulations needed to accurately predict non-Gaussian observables are computationally expensive, we reviewed in Chapter 4 what the minimal requirements would be for an LSST-like survey and for a suite of commonly used WL statistics: the power spectrum, the one-point probability density function, lensing peaks, and Minkowski functionals. We found that the total computational cost can be reduced by an order of magnitude lowering the mass resolution from a typical value of  $9 \times 10^{10} h^{-1} M_{\odot}$  to  $7 \times 10^{11} h^{-1} M_{\odot}$ . Trying to increase the number of pseudo-independent WL maps recycled from each N-body simulation reducing the thickness of the lensing planes used to build past light cones is not advisable, for below  $60 h^{-1}$  Mpc the loss of power due to the finite planes' window function affects the accuracy of non-Gaussian statistics.

In Chapter 5 we propose to use the signal imprinted in the CMB by the rotation of galaxies' ionized gaseous envelopes (rotational kinetic SZ effect) as an additional observation to constrain the baryon distribution on scales  $\lesssim 1$  Mpc. Stacked analyses of a few 100's of  $\text{deg}^2$  with complete information of galaxy spins, over CMB maps with  $\approx 1.5$  arcmin resolution, should be enough to statistically detect this effect. Constraining the baryonic matter distribution on this scales is critical to accurately model the WL signal at the  $\sim 1$  arcmin resolution reached by upcoming experiments.

In Chapter 6 we show for the first time that DNNs can be trained on simulated WL data to extract information not accessible to some statistics currently used to analyze survey data (power spectrum and lensing peaks). Under idealized conditions, DNNs can yield constraints up to approximately five times tighter than two-point statistics, providing a lower bound on the amount of non-Gaussian information encoded in WL data sets.

Finally, in Chapter 7, we prove that DNNs are not only competitive just with individual non-Gaussian statistics, but also with combinations of them. We demonstrate how back-propagation based saliency methods are not appropriate to interpret the output of neural networks trained on WL data. Gradient-based methods show how, at least for noiseless data and a high-performing DNN, pixels with extreme  $\kappa$  values are the most relevant to the network's output. On aggregate, regions with negative  $\kappa$  are the most relevant. This is a potentially encouraging result, because these regions trace cosmic voids, which are robust to systematic effects from baryonic physics,

and sensitive to modified gravity models.

## 8.2 Future work

Looking forward, it is possible to build upon several of the lines of work presented in this thesis. While DNNs have already been used in the context of survey data, in order to rely on their results, we need a transparent understanding of how their output depends on the underlying physical model of the data, and a better characterization of how their output is affected by the presence of systematics in the data. Using likelihood-free methods on the output of DNNs could serve as a fully non-parametric inference scheme from pixels to cosmological parameters. There is still room for exploring which network architectures work best with WL data sets (based on their symmetries, noise and masking properties, etc.).

The development of fast, accurate emulators for specific non-Gaussian statistics will facilitate the analysis of the measurements from future surveys. Existing models such as CAMELUS can be modified for this purpose. Ultimately, we need to address the theoretical limitations of current models predicting the spatial distribution of matter on small scales, including baryons. The work presented on the rotational kSZ effect can be extended, applying it to ongoing and upcoming CMB experiments, and using hydrodynamical simulations, instead of analytic models, to interpret the signal. Any future detection can be combined with measurements of the kinematic and thermal SZ effects, as well as measurements of the X-ray and UV emission to characterize the detailed distribution of baryons around galaxies.



## References

- [1] A. G. Riess, A. V. Filippenko, P. Challis, et al., “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant,” *AJ*, vol. 116, no. 3, pp. 1009–1038, Sep. 1998. arXiv: [astro-ph/9805201](#) [[astro-ph](#)].
- [2] S. Perlmutter, G. Aldering, G. Goldhaber, et al., “Measurements of  $\Omega$  and  $\Lambda$  from 42 High-Redshift Supernovae,” *ApJ*, vol. 517, no. 2, pp. 565–586, Jun. 1999. arXiv: [astro-ph/9812133](#) [[astro-ph](#)].
- [3] A. Albrecht, G. Bernstein, R. Cahn, et al., “Report of the Dark Energy Task Force,” *arXiv e-prints*, Sep. 2006. arXiv: [astro-ph/0609591](#) [[astro-ph](#)].
- [4] M. Bartelmann and P. Schneider, “Weak gravitational lensing,” *Phys. Rep.*, vol. 340, pp. 291–472, Jan. 2001. arXiv: [astro-ph/9912508](#) [[astro-ph](#)].
- [5] M. Kilbinger, “Cosmology with cosmic shear observations: a review,” *Rep. Prog. Phys.*, vol. 78, 086901, p. 086 901, Jul. 2015. arXiv: [1411.0115](#) [[astro-ph.CO](#)].
- [6] M. Garcia-Fernandez, E. Sánchez, I. Sevilla-Noarbe, et al., “Weak lensing magnification in the Dark Energy Survey Science Verification Data,” *arXiv e-prints*, arXiv:1611.10326, arXiv:1611.10326, Nov. 2016. arXiv: [1611.10326](#) [[astro-ph.CO](#)].
- [7] S. Dodelson, *Gravitational Lensing*. 2017.
- [8] R. Mandelbaum, “Weak Lensing for Precision Cosmology,” *ARA&A*, vol. 56, pp. 393–433, Sep. 2018. arXiv: [1710.03235](#) [[astro-ph.CO](#)].
- [9] N. Kaiser and G. Squires, “Mapping the Dark Matter with Weak Gravitational Lensing,” *ApJ*, vol. 404, p. 441, Feb. 1993.
- [10] P. J. Marshall, M. P. Hobson, S. F. Gull, and S. L. Bridle, “Maximum-entropy weak lens reconstruction: improved methods and application to data,” *MNRAS*, vol. 335, no. 4, pp. 1037–1048, Oct. 2002. arXiv: [astro-ph/0112396](#) [[astro-ph](#)].
- [11] C. Pichon, E. Thiébaud, S. Prunet, et al., “ASKI: full-sky lensing map-making algorithms,” *MNRAS*, vol. 401, no. 2, pp. 705–726, Jan. 2010. arXiv: [0901.2001](#) [[astro-ph.CO](#)].
- [12] N. Jeffrey, F. B. Abdalla, O. Lahav, et al., “Improving weak lensing mass map reconstructions using Gaussian and sparsity priors: application to DES SV,” *MNRAS*, vol. 479, no. 3, pp. 2871–2888, Sep. 2018. arXiv: [1801.08945](#) [[astro-ph.CO](#)].

- [13] S. Pires, V. Vandenbussche, V. Kansal, *et al.*, “Euclid: Reconstruction of Weak Lensing mass maps for non-Gaussianity studies,” *arXiv e-prints*, arXiv:1910.03106, arXiv:1910.03106, Oct. 2019. arXiv: [1910.03106 \[astro-ph.CO\]](#).
- [14] D. J. Bacon, A. R. Refregier, and R. S. Ellis, “Detection of weak gravitational lensing by large-scale structure,” *MNRAS*, vol. 318, no. 2, pp. 625–640, Oct. 2000. arXiv: [astro-ph/0003008 \[astro-ph\]](#).
- [15] L. Van Waerbeke, Y. Mellier, T. Erben, *et al.*, “Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures,” *A&A*, vol. 358, pp. 30–44, Jun. 2000. arXiv: [astro-ph/0002500 \[astro-ph\]](#).
- [16] N. Kaiser, G. Wilson, and G. A. Luppino, “Large-Scale Cosmic Shear Measurements,” *arXiv e-prints*, astro-ph/0003338, astro-ph/0003338, Mar. 2000. arXiv: [astro-ph/0003338 \[astro-ph\]](#).
- [17] D. M. Wittman, J. A. Tyson, D. Kirkman, I. Dell’Antonio, and G. Bernstein, “Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales,” *Nature*, vol. 405, no. 6783, pp. 143–148, May 2000. arXiv: [astro-ph/0003014 \[astro-ph\]](#).
- [18] A. Nicola, A. Refregier, and A. Amara, “Integrated approach to cosmology: Combining CMB, large-scale structure, and weak lensing,” *PhRvD*, vol. 94, no. 8, 083517, p. 083 517, Oct. 2016. arXiv: [1607.01014 \[astro-ph.CO\]](#).
- [19] M. A. Troxel, N. MacCrann, J. Zuntz, *et al.*, “Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear,” *PhRvD*, vol. 98, no. 4, 043528, p. 043 528, Aug. 2018. arXiv: [1708.01538 \[astro-ph.CO\]](#).
- [20] C. Hikage, M. Oguri, T. Hamana, *et al.*, “Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data,” *Publ. Astron. Soc. Jpn.*, vol. 71, no. 2, 43, p. 43, Apr. 2019. arXiv: [1809.09148 \[astro-ph.CO\]](#).
- [21] LSST Science Collaboration, P. A. Abell, J. Allison, *et al.*, “LSST Science Book, Version 2.0,” *arXiv e-prints*, Dec. 2009. arXiv: [0912.0201 \[astro-ph.IM\]](#).
- [22] A. Refregier, A. Amara, T. D. Kitching, *et al.*, “Euclid Imaging Consortium Science Book,” *arXiv e-prints*, Jan. 2010. arXiv: [1001.0061 \[astro-ph.IM\]](#).
- [23] J. Rhodes, D. Alonso, B. Ansarinejad, *et al.*, “Cosmological Synergies Enabled by Joint Analysis of Multi-probe data from WFIRST, Euclid, and LSST,” *Astro2020*, vol. 51, no. 3, 201, p. 201, May 2019.

- [24] D. Spergel, N. Gehrels, C. Baltay, *et al.*, “Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report,” *arXiv e-prints*, Mar. 2015. arXiv: [1503.03757 \[astro-ph.IM\]](#).
- [25] T. Erben, H. Hildebrandt, L. Miller, *et al.*, “CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey - imaging data and catalogue products,” *MNRAS*, vol. 433, no. 3, pp. 2545–2563, Aug. 2013. arXiv: [1210.8156 \[astro-ph.CO\]](#).
- [26] J. T. A. de Jong, G. A. Verdoes Kleijn, K. H. Kuijken, and E. A. Valentijn, “The Kilo-Degree Survey,” *Experimental Astronomy*, vol. 35, no. 1-2, pp. 25–44, Jan. 2013. arXiv: [1206.1254 \[astro-ph.CO\]](#).
- [27] C. Chang, M. Wang, S. Dodelson, *et al.*, “A unified analysis of four cosmic shear surveys,” *MNRAS*, vol. 482, no. 3, pp. 3696–3717, Jan. 2019. arXiv: [1808.07335 \[astro-ph.CO\]](#).
- [28] M. Takada and B. Jain, “Cosmological parameters from lensing power spectrum and bispectrum tomography,” *MNRAS*, vol. 348, no. 3, pp. 897–915, Mar. 2004. arXiv: [astro-ph/0310125 \[astro-ph\]](#).
- [29] M. Sato and T. Nishimichi, “Impact of the non-Gaussian covariance of the weak lensing power spectrum and bispectrum on cosmological parameter estimation,” *PhRvD*, vol. 87, no. 12, 123538, p. 123 538, Jun. 2013. arXiv: [1301.3588 \[astro-ph.CO\]](#).
- [30] I. Kayo, M. Takada, and B. Jain, “Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix,” *MNRAS*, vol. 429, no. 1, pp. 344–371, Feb. 2013. arXiv: [1207.6322 \[astro-ph.CO\]](#).
- [31] J. P. Dietrich and J. Hartlap, “Cosmology with the shear-peak statistics,” *MNRAS*, vol. 402, no. 2, pp. 1049–1058, Feb. 2010. arXiv: [0906.3512 \[astro-ph.CO\]](#).
- [32] J. M. Kratochvil, Z. Haiman, and M. May, “Probing cosmology with weak lensing peak counts,” *PhRvD*, vol. 81, 043519, p. 043 519, Feb. 2010. arXiv: [0907.0486 \[astro-ph.CO\]](#).
- [33] L. Marian, R. E. Smith, S. Hilbert, and P. Schneider, “The cosmological information of shear peaks: beyond the abundance,” *MNRAS*, vol. 432, no. 2, pp. 1338–1350, Jun. 2013. arXiv: [1301.5001 \[astro-ph.CO\]](#).
- [34] D. Ribli, B. Á. Pataki, and I. Csabai, “An improved cosmological parameter inference scheme motivated by deep learning,” *Nature Astronomy*, vol. 3, pp. 93–98, Jan. 2019. arXiv: [1806.05995 \[astro-ph.CO\]](#).
- [35] C. T. Davies, M. Cautun, and B. Li, “Weak lensing by voids in weak lensing maps,” *MNRAS*, vol. 480, no. 1, pp. L101–L105, Oct. 2018. arXiv: [1803.08717 \[astro-ph.CO\]](#).

- [36] E. Paillas, M. Cautun, B. Li, *et al.*, “The Santiago-Harvard-Edinburgh-Durham void comparison II: unveiling the Vainshtein screening using weak lensing,” *MNRAS*, vol. 484, no. 1, pp. 1149–1165, Mar. 2019. arXiv: [1810.02864 \[astro-ph.CO\]](#).
- [37] C. D. Kreisch, A. Pisani, C. Carbone, *et al.*, “Massive neutrinos leave fingerprints on cosmic voids,” *MNRAS*, vol. 488, no. 3, pp. 4413–4426, Sep. 2019. arXiv: [1808.07464 \[astro-ph.CO\]](#).
- [38] E. Paillas, C. D. P. Lagos, N. Padilla, *et al.*, “Baryon effects on void statistics in the EAGLE simulation,” *MNRAS*, vol. 470, no. 4, pp. 4434–4452, Oct. 2017. arXiv: [1609.00101 \[astro-ph.CO\]](#).
- [39] W. R. Coulton, J. Liu, I. G. McCarthy, and K. Osato, “Weak Lensing Minima and Peaks: Cosmological Constraints and the Impact of Baryons,” *MNRAS*, Apr. 2020. arXiv: [1910.04171 \[astro-ph.CO\]](#).
- [40] D. Munshi, L. van Waerbeke, J. Smidt, and P. Coles, “From weak lensing to non-Gaussianity via Minkowski functionals,” *MNRAS*, vol. 419, no. 1, pp. 536–555, Jan. 2012. arXiv: [1103.1876 \[astro-ph.CO\]](#).
- [41] J. M. Kratochvil, E. A. Lim, S. Wang, Z. Haiman, M. May, and K. Huffenberger, “Probing cosmology with weak lensing Minkowski functionals,” *PhRvD*, vol. 85, no. 10, 103513, p. 103 513, May 2012. arXiv: [1109.6334 \[astro-ph.CO\]](#).
- [42] A. Petri, Z. Haiman, L. Hui, M. May, and J. M. Kratochvil, “Cosmology with Minkowski functionals and moments of the weak lensing convergence field,” *PhRvD*, vol. 88, 123002, p. 123 002, Dec. 2013. arXiv: [1309.4460 \[astro-ph.CO\]](#).
- [43] C. Parroni, V. F. Cardone, R. Maoli, and R. Scaramella, “Going deep with Minkowski functionals of convergence maps,” *A&A*, vol. 633, A71, A71, Jan. 2020. arXiv: [1911.06243 \[astro-ph.CO\]](#).
- [44] G. A. Marques, J. Liu, J. M. Zorrilla Matilla, *et al.*, “Constraining neutrino mass with weak lensing Minkowski Functionals,” *JCAP*, vol. 2019, no. 6, 019, p. 019, Jun. 2019. arXiv: [1812.08206 \[astro-ph.CO\]](#).
- [45] L. Fu, M. Kilbinger, T. Erben, *et al.*, “CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations,” *MNRAS*, vol. 441, no. 3, pp. 2725–2743, Jul. 2014. arXiv: [1404.5469 \[astro-ph.CO\]](#).
- [46] H. Y. Shan, J.-P. Kneib, J. Comparat, E. Jullo, A. Charbonnier, T. Erben, M. Makler, B. Moraes, L. Van Waerbeke, F. Courbin, G. Meylan, C. Tao, and J. E. Taylor, “Weak lensing mass map and peak statistics in Canada-France-Hawaii Telescope Stripe 82 survey,” *MNRAS*, vol. 442, no. 3, pp. 2534–2542, Aug. 2014. arXiv: [1311.1319 \[astro-ph.CO\]](#).

- [47] J. Liu, A. Petri, Z. Haiman, et al., “Cosmology constraints from the weak lensing peak counts and the power spectrum in CFHTLenS data,” *PhRvD*, vol. 91, no. 6, 063507, p. 063 507, Mar. 2015. arXiv: [1412.0757 \[astro-ph.CO\]](#).
- [48] X. Liu, C. Pan, R. Li, et al., “Cosmological constraints from weak lensing peak statistics with Canada-France-Hawaii Telescope Stripe 82 Survey,” *MNRAS*, vol. 450, no. 3, pp. 2888–2902, Jul. 2015. arXiv: [1412.3683 \[astro-ph.CO\]](#).
- [49] M. Shirasaki and N. Yoshida, “Statistical and Systematic Errors in the Measurement of Weak-Lensing Minkowski Functionals: Application to the Canada-France-Hawaii Lensing Survey,” *ApJ*, vol. 786, no. 1, 43, p. 43, May 2014. arXiv: [1312.5032 \[astro-ph.CO\]](#).
- [50] A. Petri, J. Liu, Z. Haiman, et al., “Emulating the CFHTLenS weak lensing data: Cosmological constraints from moments and Minkowski functionals,” *PhRvD*, vol. 91, no. 10, 103511, p. 103 511, May 2015. arXiv: [1503.06214 \[astro-ph.CO\]](#).
- [51] H. Shan, X. Liu, H. Hildebrandt, et al., “KiDS-450: cosmological constraints from weak lensing peak statistics - I. Inference from analytical prediction of high signal-to-noise ratio convergence peaks,” *MNRAS*, vol. 474, no. 1, pp. 1116–1134, Feb. 2018. arXiv: [1709.07651 \[astro-ph.CO\]](#).
- [52] N. Martinet, P. Schneider, H. Hildebrandt, et al., “KiDS-450: cosmological constraints from weak-lensing peak statistics - II: Inference from shear peaks using N-body simulations,” *MNRAS*, vol. 474, no. 1, pp. 712–730, Feb. 2018. arXiv: [1709.07678 \[astro-ph.CO\]](#).
- [53] T. Kacprzak, D. Kirk, O. Friedrich, A. Amara, et al., “Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data,” *MNRAS*, vol. 463, no. 4, pp. 3653–3673, Dec. 2016. arXiv: [1603.05040 \[astro-ph.CO\]](#).
- [54] D. Gruen, O. Friedrich, E. Krause, et al., “Density split statistics: Cosmological constraints from counts and lensing in cells in DES Y1 and SDSS data,” *PhRvD*, vol. 98, no. 2, 023507, p. 023 507, Jul. 2018. arXiv: [1710.05045 \[astro-ph.CO\]](#).
- [55] T. Hamana, J. Sakurai, M. Koike, and L. Miller, “Cosmological constraints from Subaru weak lensing cluster counts,” *Publ. Astron. Soc. Jpn.*, vol. 67, no. 3, 34, p. 34, Jun. 2015. arXiv: [1503.01851 \[astro-ph.CO\]](#).
- [56] S. Miyazaki, M. Oguri, T. Hamana, et al., “A large sample of shear-selected clusters from the Hyper Suprime-Cam Subaru Strategic Program S16A Wide field mass maps,” *Publ. Astron. Soc. Jpn.*, vol. 70, S27, S27, Jan. 2018. arXiv: [1802.10290 \[astro-ph.CO\]](#).
- [57] W. McCulloch and W. Pitts, “A logical calculus of ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, vol. 5, pp. 127–147, 1943.

- [58] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” Psychological Review, pp. 65–386, 1958.
- [59] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012, pp. 1097–1105.
- [60] D. Silver, A. Huang, C. J. Maddison, A. Guez, et al., “Mastering the game of Go with deep neural networks and tree search,” Nature, vol. 529, no. 7587, pp. 484–489, Jan. 2016.
- [61] N. Brown and T. Sandholm, “Superhuman AI for heads-up no-limit poker: Libratus beats top professionals,” Science, vol. 359, no. 6374, pp. 418–424, Jan. 2018.
- [62] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
- [63] S. Sengupta, S. Basak, P. Saikia, et al., “A Review of Deep Learning with Special Emphasis on Architectures, Applications and Recent Trends,” arXiv e-prints, arXiv:1905.13294, arXiv:1905.13294, May 2019. arXiv: [1905.13294 \[cs.LG\]](#).
- [64] T. Serre, “Deep learning: The good, the bad, and the ugly,” Annual Review of Vision Science, vol. 5, no. 1, pp. 399–426, 2019, PMID: 31394043. eprint: <https://doi.org/10.1146/annurev-vision-091718-014951>.
- [65] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, “An introductory review of deep learning for prediction models with big data,” Frontiers in Artificial Intelligence, vol. 3, p. 4, 2020.
- [66] M. Abadi, A. Agarwal, P. Barham, et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” arXiv e-prints, arXiv:1603.04467, arXiv:1603.04467, Mar. 2016. arXiv: [1603.04467 \[cs.DC\]](#).
- [67] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” Mathematics of control, signals and systems, vol. 2, no. 4, pp. 303–314, 1989.
- [68] C. K. Chui, X. Li, and H. N. Mhaskar, “Limitations of the approximation capabilities of neural networks with one hidden layer,” Advances in Computational Mathematics, vol. 5, no. 1, pp. 233–243, 1996.
- [69] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, “The expressive power of neural networks: A view from the width,” in Advances in Neural Information Processing Systems 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017, pp. 6231–6239.

- [70] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” Nature, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
- [71] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.
- [72] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in Proc. ICML, vol. 30, 2013.
- [73] P. Simard, Y. LeCun, and D. J., “Efficient pattern recognition using a new transformation distance,” in Advances in Neural Information Processing Systems (NIPS 1992), S. Hanson, J. Cowan, and L. Giles, Eds., vol. 5, Morgan Kaufmann, 1993.
- [74] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” Neural Computation, vol. 1, no. 4, pp. 541–551, 1989.
- [75] T. S. Cohen, M. Geiger, J. Koehler, and M. Welling, “Spherical CNNs,” arXiv e-prints, arXiv:1801.10130, arXiv:1801.10130, Jan. 2018. arXiv: [1801.10130 \[cs.LG\]](#).
- [76] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, “DeepSphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications,” Astronomy and Computing, vol. 27, 130, p. 130, Apr. 2019. arXiv: [1810.12186 \[astro-ph.CO\]](#).
- [77] T. S. Cohen and M. Welling, “Group Equivariant Convolutional Networks,” arXiv e-prints, arXiv:1602.07576, arXiv:1602.07576, Feb. 2016. arXiv: [1602.07576 \[cs.LG\]](#).
- [78] M. Belkin, D. Hsu, S. Ma, and S. Mandal, “Reconciling modern machine learning practice and the bias-variance trade-off,” arXiv e-prints, arXiv:1812.11118, arXiv:1812.11118, Dec. 2018. arXiv: [1812.11118 \[stat.ML\]](#).
- [79] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, “Understanding deep learning requires rethinking generalization,” arXiv e-prints, arXiv:1611.03530, arXiv:1611.03530, Nov. 2016. arXiv: [1611.03530 \[cs.LG\]](#).
- [80] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K. Müller, Eds., Explainable AI: Interpreting, Explaining, and Predicting to Make AI Decisions. Lecture Notes in Computer Science. Springer, 2019, vol. 11700, ISBN: 978-3-030-28953-9.
- [81] M. Tulio Ribeiro, S. Singh, and C. Guestrin, ““Why Should I Trust You?”: Explaining the Predictions of Any Classifier,” arXiv e-prints, arXiv:1602.04938, arXiv:1602.04938, Feb. 2016. arXiv: [1602.04938 \[cs.LG\]](#).



- [82] R. Fong and A. Vedaldi, “Interpretable Explanations of Black Boxes by Meaningful Perturbation,” arXiv e-prints, arXiv:1704.03296, arXiv:1704.03296, Apr. 2017. arXiv: [1704.03296 \[cs.CV\]](#).
- [83] L. M. Zintgraf, T. S. Cohen, T. Adel, and M. Welling, “Visualizing Deep Neural Network Decisions: Prediction Difference Analysis,” arXiv e-prints, arXiv:1702.04595, arXiv:1702.04595, Feb. 2017. arXiv: [1702.04595 \[cs.CV\]](#).
- [84] A. Nguyen, J. Yosinski, and J. Clune, “Understanding Neural Networks via Feature Visualization: A survey,” arXiv e-prints, arXiv:1904.08939, arXiv:1904.08939, Apr. 2019. arXiv: [1904.08939 \[cs.LG\]](#).
- [85] M. Ancona, E. Ceolini, C. Öztireli, and M. Gross, “Towards better understanding of gradient-based attribution methods for Deep Neural Networks,” arXiv e-prints, arXiv:1711.06104, arXiv:1711.06104, Nov. 2017. arXiv: [1711.06104 \[cs.LG\]](#).
- [86] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, “On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation,” PLoS ONE, vol. 10, no. 7, e0130140, Jul. 2015.
- [87] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” arXiv e-prints, arXiv:1311.2901, arXiv:1311.2901, Nov. 2013. arXiv: [1311.2901 \[cs.CV\]](#).
- [88] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for Simplicity: The All Convolutional Net,” arXiv e-prints, arXiv:1412.6806, arXiv:1412.6806, Dec. 2014. arXiv: [1412.6806 \[cs.LG\]](#).
- [89] J. Adebayo, J. Gilmer, M. Muelly, et al., “Sanity Checks for Saliency Maps,” arXiv e-prints, arXiv:1810.03292, arXiv:1810.03292, Oct. 2018. arXiv: [1810.03292 \[cs.CV\]](#).
- [90] M. Ntampaka, C. Avestruz, S. Boada, et al., “The Role of Machine Learning in the Next Decade of Cosmology,” Astro2020, vol. 51, no. 3, 14, p. 14, May 2019. arXiv: [1902.10159 \[astro-ph.IM\]](#).
- [91] C. J. Fluke and C. Jacobs, “Surveying the reach and maturity of machine learning and artificial intelligence in astronomy,” WIREs Data Mining and Knowledge Discovery, vol. 10, no. 2, e1349, e1349, Jan. 2020. arXiv: [1912.02934 \[astro-ph.IM\]](#).
- [92] J. Herbel, T. Kacprzak, A. Amara, A. Refregier, and A. Lucchi, “Fast point spread function modeling with deep learning,” JCAP, vol. 2018, no. 7, 054, p. 054, Jul. 2018. arXiv: [1801.07615 \[astro-ph.IM\]](#).
- [93] D. Ribli, L. Dobos, and I. Csabai, “Galaxy shape measurement with convolutional neural networks,” MNRAS, vol. 489, no. 4, pp. 4847–4859, Nov. 2019. arXiv: [1902.08161 \[astro-ph.CO\]](#).



- [94] D. Gruen, S. Seitz, J. Koppenhoefer, and A. Riffeser, “Bias-free Shear Estimation Using Artificial Neural Networks,” *ApJ*, vol. 720, no. 1, pp. 639–651, Sep. 2010. arXiv: [1002.0838 \[astro-ph.CO\]](#).
- [95] M. Tewes, T. Kuntzer, R. Nakajima, et al., “Weak-lensing shear measurement with machine learning. Teaching artificial neural networks about feature noise,” *A&A*, vol. 621, A36, A36, Jan. 2019. arXiv: [1807.02120 \[astro-ph.CO\]](#).
- [96] O. M. Springer, E. O. Ofek, Y. Weiss, and J. Merten, “Weak lensing shear estimation beyond the shape-noise limit: a machine learning approach,” *MNRAS*, vol. 491, no. 4, pp. 5301–5316, Feb. 2020. arXiv: [1808.07491 \[astro-ph.CO\]](#).
- [97] A. E. Firth, O. Lahav, and R. S. Somerville, “Estimating photometric redshifts with artificial neural networks,” *MNRAS*, vol. 339, no. 4, pp. 1195–1202, Mar. 2003. arXiv: [astro-ph/0203250 \[astro-ph\]](#).
- [98] C. Bonnett, “Using neural networks to estimate redshift distributions. An application to CFHTLenS,” *MNRAS*, vol. 449, no. 1, pp. 1043–1056, May 2015. arXiv: [1312.1287 \[astro-ph.CO\]](#).
- [99] B. Hoyle, “Measuring photometric redshifts using galaxy images and Deep Neural Networks,” *Astronomy and Computing*, vol. 16, pp. 34–40, Jul. 2016. arXiv: [1504.07255 \[astro-ph.IM\]](#).
- [100] M. Bilicki, H. Hoekstra, M. J. I. Brown, et al., “Photometric redshifts for the Kilo-Degree Survey. Machine-learning analysis with artificial neural networks,” *A&A*, vol. 616, A69, A69, Aug. 2018. arXiv: [1709.04205 \[astro-ph.CO\]](#).
- [101] A. D’Isanto and K. L. Polsterer, “Photometric redshift estimation via deep learning. Generalized and pre-classification-less, image based, fully probabilistic redshifts,” *A&A*, vol. 609, A111, A111, Jan. 2018. arXiv: [1706.02467 \[astro-ph.IM\]](#).
- [102] J. Pasquet, E. Bertin, M. Treyer, S. Arnouts, and D. Fouchez, “Photometric redshifts from SDSS images using a convolutional neural network,” *A&A*, vol. 621, A26, A26, Jan. 2019. arXiv: [1806.06607 \[astro-ph.IM\]](#).
- [103] M. R. I. Syarifudin, M. I. Hakim, and M. I. Arifyanto, “Applying Deep Neural Networks (DNN) for Measuring Photometric Redshifts from Galaxy Images: Preliminary Study,” in *Journal of Physics Conference Series*, ser. Journal of Physics Conference Series, vol. 1231, May 2019, p. 012 013.
- [104] M. Eriksen, A. Alarcon, L. Cabayol, et al., “The PAU Survey: Photometric redshifts using transfer learning from simulations,” *arXiv e-prints*, arXiv:2004.07979, arXiv:2004.07979, Apr. 2020. arXiv: [2004.07979 \[astro-ph.GA\]](#).

- [105] J.-E. Campagne, “Adversarial training applied to Convolutional Neural Network for photometric redshift predictions,” arXiv e-prints, arXiv:2002.10154, arXiv:2002.10154, Feb. 2020. arXiv: [2002.10154 \[astro-ph.IM\]](#).
- [106] N. Jeffrey, F. Lanusse, O. Lahav, and J.-L. Starck, “Deep learning dark matter map reconstructions from DES SV weak lensing data,” MNRAS, vol. 492, no. 4, pp. 5023–5029, Mar. 2020. arXiv: [1908.00543 \[astro-ph.CO\]](#).
- [107] I. Goodfellow, “NIPS 2016 Tutorial: Generative Adversarial Networks,” arXiv e-prints, arXiv:1701.00160, arXiv:1701.00160, Dec. 2016. arXiv: [1701.00160 \[cs.LG\]](#).
- [108] M. Shirasaki, N. Yoshida, and S. Ikeda, “Denoising weak lensing mass maps with deep learning,” PhRvD, vol. 100, no. 4, 043527, p. 043 527, Aug. 2019. arXiv: [1812.05781 \[astro-ph.CO\]](#).
- [109] M. Shirasaki, N. Yoshida, S. Ikeda, T. Oogi, and T. Nishimichi, “Decoding Cosmological Information in Weak-Lensing Mass Maps with Generative Adversarial Networks,” arXiv e-prints, arXiv:1911.12890, arXiv:1911.12890, Nov. 2019. arXiv: [1911.12890 \[astro-ph.CO\]](#).
- [110] A. C. Rodríguez, T. Kacprzak, A. Lucchi, et al., “Fast cosmic web simulations with generative adversarial networks,” CompAC, vol. 5, no. 1, 4, p. 4, Nov. 2018. arXiv: [1801.09070 \[astro-ph.CO\]](#).
- [111] M. Mustafa, D. Bard, W. Bhimji, et al., “CosmoGAN: creating high-fidelity weak lensing convergence maps using Generative Adversarial Networks,” CompAC, vol. 6, no. 1, 1, p. 1, May 2019. arXiv: [1706.02390 \[astro-ph.IM\]](#).
- [112] N. Perraudin, S. Marcon, A. Lucchi, and T. Kacprzak, “Emulation of cosmological mass maps with conditional generative adversarial networks,” arXiv e-prints, arXiv:2004.08139, arXiv:2004.08139, Apr. 2020. arXiv: [2004.08139 \[astro-ph.CO\]](#).
- [113] J. Schmelzle, A. Lucchi, T. Kacprzak, et al., “Cosmological model discrimination with Deep Learning,” arXiv e-prints, arXiv:1707.05167, arXiv:1707.05167, Jul. 2017. arXiv: [1707.05167 \[astro-ph.CO\]](#).
- [114] A. Gupta, J. M. Z. Matilla, D. Hsu, and Z. Haiman, “Non-Gaussian information from weak lensing data via deep learning,” PhRvD, vol. 97, no. 10, 103515, p. 103 515, May 2018. arXiv: [1802.01212 \[astro-ph.CO\]](#).
- [115] J. Fluri, T. Kacprzak, A. Refregier, A. Amara, A. Lucchi, and T. Hofmann, “Cosmological constraints from noisy convergence maps through deep learning,” PhRvD, vol. 98, no. 12, 123518, p. 123 518, Dec. 2018. arXiv: [1807.08732 \[astro-ph.CO\]](#).

- [116] D. Ribli, B. Á. Pataki, J. M. Zorrilla Matilla, D. Hsu, Z. Haiman, and I. Csabai, “Weak lensing cosmology with convolutional neural networks on noisy data,” *MNRAS*, vol. 490, no. 2, pp. 1843–1860, Dec. 2019. arXiv: [1902.03663 \[astro-ph.CO\]](#).
- [117] J. Fluri, T. Kacprzak, A. Lucchi, et al., “Cosmological constraints with deep learning from KiDS-450 weak lensing maps,” *PhRvD*, vol. 100, no. 6, 063514, p. 063 514, Sep. 2019. arXiv: [1906.03156 \[astro-ph.CO\]](#).
- [118] J. Merten, C. Giocoli, M. Baldi, et al., “On the dissection of degenerate cosmologies with machine learning,” *MNRAS*, vol. 487, no. 1, pp. 104–122, Jul. 2019. arXiv: [1810.11027 \[astro-ph.CO\]](#).
- [119] A. Peel, F. Lalande, J.-L. Starck, et al., “Distinguishing standard and modified gravity cosmologies with machine learning,” *PhRvD*, vol. 100, no. 2, 023508, p. 023 508, Jul. 2019. arXiv: [1810.11030 \[astro-ph.CO\]](#).
- [120] J. M. Z. Matilla, Z. Haiman, A. Petri, and T. Namikawa, “Geometry and growth contributions to cosmic shear observables,” *PhRvD*, vol. 96, no. 2, 023513, p. 023 513, Jul. 2017.
- [121] J. M. Zorrilla Matilla, Z. Haiman, D. Hsu, A. Gupta, and A. Petri, “Do dark matter halos explain lensing peaks?” *PhRvD*, vol. 94, 083506, p. 083 506, Oct. 2016. arXiv: [1609.03973 \[astro-ph.CO\]](#).
- [122] J. M. Z. Matilla, S. Waterval, and Z. Haiman, “Optimizing simulation parameters for weak lensing analyses involving non-gaussian observables,” *The Astronomical Journal*, vol. 159, no. 6, p. 284, Jun. 2020.
- [123] J. M. Z. Matilla and Z. Haiman, “Probing gaseous galactic halos through the rotational kinematic Sunyaev-Zeldovich effect,” *PhRvD*, vol. 101, no. 8, 083016, p. 083 016, Apr. 2020.
- [124] L. Anderson, É. Aubourg, S. Bailey, et al., “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples,” *MNRAS*, vol. 441, no. 1, pp. 24–62, Jun. 2014. arXiv: [1312.4877 \[astro-ph.CO\]](#).
- [125] R. A. Alpher, H. Bethe, and G. Gamow, “The Origin of Chemical Elements,” *Physical Review*, vol. 73, no. 7, pp. 803–804, Apr. 1948.
- [126] C. J. Copi, D. N. Schramm, and M. S. Turner, “Big-Bang Nucleosynthesis and the Baryon Density of the Universe,” *Science*, vol. 267, no. 5195, pp. 192–199, Jan. 1995. arXiv: [astro-ph/9407006 \[astro-ph\]](#).
- [127] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al., “Planck 2015 results. XIII. Cosmological parameters,” *A&A*, vol. 594, A13, A13, Sep. 2016.

- [128] M. Fukugita and P. J. E. Peebles, “The Cosmic Energy Inventory,” *ApJ*, vol. 616, no. 2, pp. 643–668, Dec. 2004. arXiv: [astro-ph/0406095](#) [[astro-ph](#)].
- [129] F. Bernardeau, L. van Waerbeke, and Y. Mellier, “Weak lensing statistics as a probe of  $\{\Omega_{\text{M}}\}$  and power spectrum,” *A&A*, vol. 322, pp. 1–18, Jun. 1997. arXiv: [astro-ph/9609122](#) [[astro-ph](#)].
- [130] L. Hui, “Weighing the Cosmological Energy Contents with Weak Gravitational Lensing,” *ApJL*, vol. 519, no. 1, pp. L9–L12, Jul. 1999. arXiv: [astro-ph/9902275](#) [[astro-ph](#)].
- [131] D. Munshi and B. Jain, “Statistics of weak lensing at small angular scales: analytical predictions for lower order moments,” *MNRAS*, vol. 322, no. 1, pp. 107–120, Mar. 2001. arXiv: [astro-ph/9912330](#) [[astro-ph](#)].
- [132] B. Jain and L. Van Waerbeke, “Statistics of Dark Matter Halos from Gravitational Lensing,” *ApJL*, vol. 530, no. 1, pp. L1–L4, Feb. 2000. arXiv: [astro-ph/9910459](#) [[astro-ph](#)].
- [133] J. Sato, M. Takada, Y. P. Jing, and T. Futamase, “Implication of  $\Omega_{\text{M}}$  through the Morphological Analysis of Weak Lensing Fields,” *ApJL*, vol. 551, no. 1, pp. L5–L8, Apr. 2001. arXiv: [astro-ph/0104015](#) [[astro-ph](#)].
- [134] F. Simpson and S. Bridle, “Illuminating dark energy with cosmic shear,” *PhRvD*, vol. 71, no. 8, 083501, p. 083 501, Apr. 2005. arXiv: [astro-ph/0411673](#) [[astro-ph](#)].
- [135] H. Zhan, L. Knox, and J. A. Tyson, “Distance, Growth Factor, and Dark Energy Constraints from Photometric Baryon Acoustic Oscillation and Weak Lensing Measurements,” *ApJ*, vol. 690, no. 1, pp. 923–936, Jan. 2009. arXiv: [0806.0937](#) [[astro-ph](#)].
- [136] B. Jain and A. Taylor, “Cross-Correlation Tomography: Measuring Dark Energy Evolution with Weak Lensing,” *PhRv*, vol. 91, no. 14, 141302, p. 141 302, Oct. 2003. arXiv: [astro-ph/0306046](#) [[astro-ph](#)].
- [137] J. Zhang, L. Hui, and A. Stebbins, “Isolating Geometry in Weak-Lensing Measurements,” *ApJ*, vol. 635, no. 2, pp. 806–820, Dec. 2005. arXiv: [astro-ph/0312348](#) [[astro-ph](#)].
- [138] S. Wang, L. Hui, M. May, and Z. Haiman, “Is modified gravity required by observations? An empirical consistency test of dark energy models,” *PhRvD*, vol. 76, no. 6, 063503, p. 063 503, Sep. 2007. arXiv: [0705.0165](#) [[astro-ph](#)].
- [139] E. J. Ruiz and D. Huterer, “Testing the dark energy consistency with geometry and growth,” *PhRvD*, vol. 91, no. 6, 063009, p. 063 009, Mar. 2015. arXiv: [1410.5832](#) [[astro-ph.CO](#)].
- [140] E. Krause and C. M. Hirata, “Weak lensing power spectra for precision cosmology. Multiple-deflection, reduced shear, and lensing bias corrections,” *A&A*, vol. 523, A28, A28, Nov. 2010. arXiv: [0910.3786](#) [[astro-ph.CO](#)].

- [141] S. Dodelson and P. Zhang, “Weak lensing bispectrum,” *PhRvD*, vol. 72, no. 8, 083001, p. 083 001, Oct. 2005. arXiv: [astro-ph/0501063](#) [[astro-ph](#)].
- [142] A. Petri, Z. Haiman, and M. May, “Validity of the Born approximation for beyond Gaussian weak lensing observables,” *PhRvD*, vol. 95, 123503, p. 123 503, Jun. 2017. arXiv: [1612.00852](#) [[astro-ph.CO](#)].
- [143] A. Petri, “Mocking the weak lensing universe: The LensTools Python computing package,” *Astronomy and Computing*, vol. 17, pp. 73–79, Oct. 2016.
- [144] V. Springel, “The cosmological simulation code GADGET-2,” *MNRAS*, vol. 364, pp. 1105–1134, Dec. 2005. eprint: [astro-ph/0505010](#).
- [145] A. Petri, Z. Haiman, and M. May, “Sample variance in weak lensing: How many simulations are required?” *PhRvD*, vol. 93, no. 6, 063524, p. 063 524, Mar. 2016. arXiv: [1601.06792](#).
- [146] R. E. Smith, J. A. Peacock, A. Jenkins, and et al., “Stable clustering, the halo model and non-linear cosmological power spectra,” *MNRAS*, vol. 341, no. 4, pp. 1311–1332, Jun. 2003. arXiv: [astro-ph/0207664](#) [[astro-ph](#)].
- [147] L. van Waerbeke, “Noise properties of gravitational lens mass reconstruction,” *MNRAS*, vol. 313, no. 3, pp. 524–532, Apr. 2000. arXiv: [astro-ph/9909160](#) [[astro-ph](#)].
- [148] P. Schneider, L. van Waerbeke, B. Jain, and G. Kruse, “A new measure for cosmic shear,” *MNRAS*, vol. 296, no. 4, pp. 873–892, Jun. 1998. arXiv: [astro-ph/9708143](#) [[astro-ph](#)].
- [149] F. Bernardeau, Y. Mellier, and L. van Waerbeke, “Detection of non-Gaussian signatures in the VIRMOS-DESCART lensing survey,” *A&A*, vol. 389, pp. L28–L32, Jul. 2002. arXiv: [astro-ph/0201032](#) [[astro-ph](#)].
- [150] K. R. Mecke, T. Buchert, and H. Wagner, “Robust morphological measures for large-scale structure in the Universe,” *A&A*, vol. 288, pp. 697–704, Aug. 1994. arXiv: [astro-ph/9312028](#) [[astro-ph](#)].
- [151] J. Hartlap, P. Simon, and P. Schneider, “Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix,” *A&A*, vol. 464, no. 1, pp. 399–404, Mar. 2007. arXiv: [astro-ph/0608064](#) [[astro-ph](#)].
- [152] P. Creminelli, L. Senatore, and M. Zaldarriaga, “Estimators for local non-Gaussianities,” *JCAP*, vol. 2007, no. 3, 019, p. 019, Mar. 2007. arXiv: [astro-ph/0606001](#) [[astro-ph](#)].
- [153] M. Tegmark, A. N. Taylor, and A. F. Heavens, “Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?” *ApJ*, vol. 480, no. 1, pp. 22–35, May 1997. arXiv: [astro-ph/9603021](#) [[astro-ph](#)].

- [154] C.-A. Lin and M. Kilbinger, “A new model to predict weak-lensing peak counts. II. Parameter constraint strategies,” *A&A*, vol. 583, A70, A70, Nov. 2015. arXiv: [1506.01076 \[astro-ph.CO\]](#).
- [155] W. Hu, “Power Spectrum Tomography with Weak Lensing,” *ApJL*, vol. 522, no. 1, pp. L21–L24, Sep. 1999. arXiv: [astro-ph/9904153 \[astro-ph\]](#).
- [156] M. White, L. van Waerbeke, and J. Mackey, “Completeness in Weak-Lensing Searches for Clusters,” *ApJ*, vol. 575, no. 2, pp. 640–649, Aug. 2002. arXiv: [astro-ph/0111490 \[astro-ph\]](#).
- [157] T. Hamana, M. Takada, and N. Yoshida, “Searching for massive clusters in weak lensing surveys,” *MNRAS*, vol. 350, no. 3, pp. 893–913, May 2004. arXiv: [astro-ph/0310607 \[astro-ph\]](#).
- [158] J. F. Hennawi and D. N. Spergel, “Shear-selected Cluster Cosmology: Tomography and Optimal Filtering,” *ApJ*, vol. 624, no. 1, pp. 59–79, May 2005. arXiv: [astro-ph/0404349 \[astro-ph\]](#).
- [159] X. Yang, J. M. Kratochvil, S. Wang, E. A. Lim, Z. Haiman, and M. May, “Cosmological information in weak lensing peaks,” *PhRvD*, vol. 84, 043529, p. 043 529, Aug. 2011. arXiv: [1109.6333 \[astro-ph.CO\]](#).
- [160] J. Liu and Z. Haiman, “Origin of weak lensing convergence peaks,” *PhRvD*, vol. 94, no. 4, 043533, p. 043 533, Aug. 2016. arXiv: [1606.01318 \[astro-ph.CO\]](#).
- [161] A. Refregier, “Weak Gravitational Lensing by Large-Scale Structure,” *ARA&A*, vol. 41, pp. 645–668, Jan. 2003. arXiv: [astro-ph/0307212 \[astro-ph\]](#).
- [162] M. Takada and B. Jain, “The three-point correlation function in cosmology,” *MNRAS*, vol. 340, no. 2, pp. 580–608, Apr. 2003. arXiv: [astro-ph/0209167 \[astro-ph\]](#).
- [163] P. Schneider, M. Kilbinger, and M. Lombardi, “The three-point correlation function of cosmic shear. II. Relation to the bispectrum of the projected mass density and generalized third-order aperture measures,” *A&A*, vol. 431, pp. 9–25, Feb. 2005. arXiv: [astro-ph/0308328 \[astro-ph\]](#).
- [164] L. Marian, R. E. Smith, S. Hilbert, and P. Schneider, “Optimized detection of shear peaks in weak lensing maps,” *MNRAS*, vol. 423, no. 2, pp. 1711–1725, Jun. 2012. arXiv: [1110.4635 \[astro-ph.CO\]](#).
- [165] J. Kwan, C. Sánchez, J. Clampitt, et al., “Cosmology from large-scale galaxy clustering and galaxy-galaxy lensing with Dark Energy Survey Science Verification data,” *MNRAS*, vol. 464, no. 4, pp. 4045–4062, Feb. 2017. arXiv: [1604.07871 \[astro-ph.CO\]](#).

- [166] Z. Fan, H. Shan, and J. Liu, “Noisy Weak-lensing Convergence Peak Statistics Near Clusters of Galaxies and Beyond,” *ApJ*, vol. 719, no. 2, pp. 1408–1420, Aug. 2010. arXiv: [1006.5121 \[astro-ph.CO\]](#).
- [167] M. Maturi, C. Angrick, F. Pace, and M. Bartelmann, “An analytic approach to number counts of weak-lensing peak detections,” *A&A*, vol. 519, A23, A23, Sep. 2010. arXiv: [0907.1849 \[astro-ph.CO\]](#).
- [168] R. Reischke, M. Maturi, and M. Bartelmann, “Extreme value statistics of weak lensing shear peak counts,” *MNRAS*, vol. 456, no. 1, pp. 641–653, Feb. 2016. arXiv: [1507.01953 \[astro-ph.CO\]](#).
- [169] K. Kainulainen and V. Marra, “Accurate modeling of weak lensing with the stochastic gravitational lensing method,” *PhRvD*, vol. 83, no. 2, 023009, p. 023 009, Jan. 2011. arXiv: [1011.0732 \[astro-ph.CO\]](#).
- [170] C.-A. Lin and M. Kilbinger, “A new model to predict weak-lensing peak counts. I. Comparison with N-body simulations,” *A&A*, vol. 576, A24, A24, Apr. 2015. arXiv: [1410.6955 \[astro-ph.CO\]](#).
- [171] G. Hinshaw, D. Larson, E. Komatsu, et al., “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results,” *ApJS*, vol. 208, no. 2, 19, p. 19, Oct. 2013. arXiv: [1212.5226 \[astro-ph.CO\]](#).
- [172] A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida, “The mass function of dark matter haloes,” *MNRAS*, vol. 321, no. 2, pp. 372–384, Feb. 2001. arXiv: [astro-ph/0005260 \[astro-ph\]](#).
- [173] J. F. Navarro, C. S. Frenk, and S. D. M. White, “A Universal Density Profile from Hierarchical Clustering,” *ApJ*, vol. 490, no. 2, pp. 493–508, Dec. 1997. arXiv: [astro-ph/9611107 \[astro-ph\]](#).
- [174] E. Sellentin and A. F. Heavens, “Parameter inference with estimated covariance matrices,” *MNRAS*, vol. 456, no. 1, pp. L132–L136, Feb. 2016. arXiv: [1511.05969 \[astro-ph.CO\]](#).
- [175] S. R. Knollmann and A. Knebe, “AHF: Amiga’s Halo Finder,” *ApJS*, vol. 182, no. 2, pp. 608–624, Jun. 2009. arXiv: [0904.3662 \[astro-ph.CO\]](#).
- [176] W. Hu and A. V. Kravtsov, “Sample Variance Considerations for Cluster Surveys,” *ApJ*, vol. 584, no. 2, pp. 702–715, Feb. 2003. arXiv: [astro-ph/0203169 \[astro-ph\]](#).
- [177] X. Yang, J. M. Kratochvil, K. Huffenberger, Z. Haiman, and M. May, “Baryon impact on weak lensing peaks and power spectrum: Low-bias statistics and self-calibration in future surveys,” *PhRvD*, vol. 87, no. 2, 023511, p. 023 511, Jan. 2013. arXiv: [1210.0608 \[astro-ph.CO\]](#).



- [178] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, et al., “Observational probes of cosmic acceleration,” Phys. Rep., vol. 530, no. 2, pp. 87–255, Sep. 2013. arXiv: [1201.2434 \[astro-ph.CO\]](#).
- [179] H. Hoekstra and B. Jain, “Weak Gravitational Lensing and Its Cosmological Applications,” Annual Review of Nuclear and Particle Science, vol. 58, no. 1, pp. 99–123, Nov. 2008. arXiv: [0805.0139 \[astro-ph\]](#).
- [180] DESI Collaboration, A. Aghamousa, J. Aguilar, et al., “The DESI Experiment Part I: Science, Targeting, and Survey Design,” arXiv e-prints, Oct. 2016. arXiv: [1611.00036 \[astro-ph.IM\]](#).
- [181] P. Bull, S. Camera, K. Kelley, et al., “Fundamental Physics with the Square Kilometer Array,” arXiv e-prints, Oct. 2018. arXiv: [1810.02680 \[astro-ph.CO\]](#).
- [182] P. Schneider and A. Weiss, “Light propagation in inhomogeneous universes - The ray-shooting method,” ApJ, vol. 330, pp. 1–15, Jul. 1988.
- [183] B. Jain, U. Seljak, and S. White, “Ray-tracing Simulations of Weak Lensing by Large-Scale Structure,” ApJ, vol. 530, pp. 547–577, Feb. 2000. eprint: [astro-ph/9901191](#).
- [184] C. Vale and M. White, “Simulating Weak Lensing by Large-Scale Structure,” ApJ, vol. 592, pp. 699–709, Aug. 2003. arXiv: [astro-ph/0303555 \[astro-ph\]](#).
- [185] S. Hilbert, J. Hartlap, S. D. M. White, and P. Schneider, “Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing,” A&A, vol. 499, pp. 31–43, May 2009. arXiv: [0809.5035 \[astro-ph\]](#).
- [186] K. Heitmann, M. White, C. Wagner, S. Habib, and D. Higdon, “The Coyote Universe. I. Precision Determination of the Nonlinear Matter Power Spectrum,” ApJ, vol. 715, pp. 104–121, May 2010. arXiv: [0812.1052 \[astro-ph\]](#).
- [187] A. Izard, P. Fosalba, and M. Crocce, “ICE-COLA: fast simulations for weak lensing observables,” MNRAS, vol. 473, pp. 3051–3061, Jan. 2018. arXiv: [1707.06312 \[astro-ph.CO\]](#).
- [188] C. Howlett, M. Manera, and W. J. Percival, “L-PICOLA: A parallel code for fast dark matter simulation,” Astronomy and Computing, vol. 12, pp. 109–126, Sep. 2015. arXiv: [1506.03737 \[astro-ph.CO\]](#).
- [189] Y. Feng, M.-Y. Chu, U. Seljak, and P. McDonald, “FASTPM: a new scheme for fast simulations of dark matter and haloes,” MNRAS, vol. 463, pp. 2273–2286, Dec. 2016. arXiv: [1603.00476 \[astro-ph.CO\]](#).
- [190] S. He, Y. Li, Y. Feng, et al., “Learning to Predict the Cosmological Structure Formation,” arXiv e-prints, Nov. 2018. arXiv: [1811.06533 \[astro-ph.CO\]](#).



- [191] A. Barreira, E. Krause, and F. Schmidt, “Accurate cosmic shear errors: do we need ensembles of simulations?” *JCAP*, vol. 10, 053, p. 053, Oct. 2018. arXiv: [1807.04266](#).
- [192] M. Sato, T. Hamana, R. Takahashi, *et al.*, “Simulations of Wide-Field Weak Lensing Surveys. I. Basic Statistics and Non-Gaussian Effects,” *ApJ*, vol. 701, pp. 945–954, Aug. 2009. arXiv: [0906.2237](#) [[astro-ph.CO](#)].
- [193] V. Springel, *NGenIC: Cosmological structure initial conditions*, Astrophysics Source Code Library, Feb. 2015. ascl: [1502.003](#).
- [194] A. Lewis, A. Challinor, and A. Lasenby, “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models,” *ApJ*, vol. 538, pp. 473–476, Aug. 2000. arXiv: [astro-ph/9911177](#) [[astro-ph](#)].
- [195] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, “Revising the Halofit Model for the Nonlinear Matter Power Spectrum,” *ApJ*, vol. 761, no. 2, 152, p. 152, Dec. 2012. arXiv: [1208.2701](#) [[astro-ph.CO](#)].
- [196] N. E. Chisari, D. Alonso, E. Krause, *et al.*, “Core Cosmology Library: Precision Cosmological Predictions for LSST,” *ApJS*, vol. 242, no. 1, 2, p. 2, May 2019. arXiv: [1812.05995](#) [[astro-ph.CO](#)].
- [197] H.-J. Huang, T. Eifler, R. Mandelbaum, and S. Dodelson, “Modeling baryonic physics in future weak lensing surveys,” *arXiv e-prints*, Sep. 2018. arXiv: [1809.01146](#) [[astro-ph.CO](#)].
- [198] N. Chisari, S. Codis, C. Laigle, Y. Dubois, C. Pichon, J. Devriendt, A. Slyz, L. Miller, R. Gavazzi, and K. Benabed, “Intrinsic alignments of galaxies in the Horizon-AGN cosmological hydrodynamical simulation,” *MNRAS*, vol. 454, no. 3, pp. 2736–2753, Dec. 2015. arXiv: [1507.07843](#) [[astro-ph.CO](#)].
- [199] The Dark Energy Survey Collaboration, “The Dark Energy Survey,” *arXiv e-prints*, Oct. 2005. arXiv: [astro-ph/0510346](#) [[astro-ph](#)].
- [200] C. Hikage, M. Oguri, T. Hamana, *et al.*, “Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data,” *arXiv e-prints*, arXiv:1809.09148, arXiv:1809.09148, Sep. 2018. arXiv: [1809.09148](#) [[astro-ph.CO](#)].
- [201] R. Takahashi, T. Hamana, M. Shirasaki, *et al.*, “Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments,” *ApJ*, vol. 850, no. 1, 24, p. 24, Nov. 2017. arXiv: [1706.01472](#) [[astro-ph.CO](#)].
- [202] S. Wang, Z. Haiman, and M. May, “Constraining Cosmology with High-Convergence Regions in Weak Lensing Surveys,” *ApJ*, vol. 691, no. 1, pp. 547–559, Jan. 2009. arXiv: [0809.4052](#) [[astro-ph](#)].

- [203] K. Patton, J. Blazek, K. Honscheid, et al., “Cosmological constraints from the convergence 1-point probability distribution,” MNRAS, vol. 472, no. 1, pp. 439–446, Nov. 2017. arXiv: [1611.01486 \[astro-ph.CO\]](#).
- [204] J. Liu and M. S. Madhavacheril, “Constraining neutrino mass with the tomographic weak lensing one-point probability distribution function and power spectrum,” PhRvD, vol. 99, no. 8, 083508, p. 083 508, Apr. 2019. arXiv: [1809.10747 \[astro-ph.CO\]](#).
- [205] J. R. Bond and G. Efstathiou, “The statistics of cosmic background radiation fluctuations,” MNRAS, vol. 226, pp. 655–687, Jun. 1987.
- [206] J. Schmalzing, M. Kerscher, and T. Buchert, “Minkowski Functionals in Cosmology,” in Dark Matter in the Universe, S. Bonometto, J. R. Primack, and A. Provenzale, Eds., Jan. 1996, p. 281. arXiv: [astro-ph/9508154 \[astro-ph\]](#).
- [207] C. Ling, Q. Wang, R. Li, B. Li, J. Wang, and L. Gao, “Distinguishing general relativity and  $f(R)$  gravity with the gravitational lensing Minkowski functionals,” PhRvD, vol. 92, no. 6, 064024, p. 064 024, Sep. 2015. arXiv: [1410.2734 \[astro-ph.CO\]](#).
- [208] T. Matsubara, “Statistics of smoothed cosmic fields in perturbation theory I: Formulation and useful formulas in second-order perturbation theory,” arXiv e-prints, astro-ph/0006269, astro-ph/0006269, Jun. 2000. arXiv: [astro-ph/0006269 \[astro-ph\]](#).
- [209] J. Tumlinson, M. S. Peebles, and J. K. Werk, “The Circumgalactic Medium,” ARA&A, vol. 55, no. 1, pp. 389–432, Aug. 2017. arXiv: [1709.09180 \[astro-ph.GA\]](#).
- [210] M. Peebles, P. Behroozi, R. Bordoloi, et al., “Understanding the circumgalactic medium is critical for understanding galaxy evolution,” in Astro2020, vol. 51, May 2019, p. 368. arXiv: [1903.05644 \[astro-ph.GA\]](#).
- [211] R. A. Sunyaev and Y. B. Zeldovich, “The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies,” Comments on Astrophysics and Space Physics, vol. 4, p. 173, Nov. 1972.
- [212] M. Birkinshaw, “The Sunyaev-Zel’dovich effect,” Phys. Rep., vol. 310, no. 2-3, pp. 97–195, Mar. 1999. arXiv: [astro-ph/9808050 \[astro-ph\]](#).
- [213] N. Hand, G. E. Addison, E. Aubourg, et al., “Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel’dovich Effect,” PhRv, vol. 109, no. 4, 041101, p. 041 101, Jul. 2012. arXiv: [1203.4219 \[astro-ph.CO\]](#).
- [214] F. De Bernardis, S. Aiola, E. M. Vavagiakis, et al., “Detection of the pairwise kinematic Sunyaev-Zel’dovich effect with BOSS DR11 and the Atacama Cosmology Telescope,” JCAP, vol. 2017, no. 3, 008, p. 008, Mar. 2017. arXiv: [1607.02139 \[astro-ph.CO\]](#).

- [215] Y.-C. Li, Y.-Z. Ma, M. Remazeilles, and K. Moodley, “Measurement of the pairwise kinematic Sunyaev-Zeldovich effect with Planck and BOSS data,” *PhRvD*, vol. 97, no. 2, 023514, p. 023 514, Jan. 2018. arXiv: [1710.10876 \[astro-ph.CO\]](#).
- [216] G. Lavaux, N. Afshordi, and M. J. Hudson, “First measurement of the bulk flow of nearby galaxies using the cosmic microwave background,” *MNRAS*, vol. 430, no. 3, pp. 1617–1635, Apr. 2013. arXiv: [1207.1721 \[astro-ph.CO\]](#).
- [217] E. Schaan, S. Ferraro, M. Vargas-Magaña, *et al.*, “Evidence for the kinematic Sunyaev-Zel’dovich effect with the Atacama Cosmology Telescope and velocity reconstruction from the Baryon Oscillation Spectroscopic Survey,” *PhRvD*, vol. 93, no. 8, 082002, p. 082 002, Apr. 2016. arXiv: [1510.06442 \[astro-ph.CO\]](#).
- [218] R. Adam, I. Bartalucci, G. W. Pratt, *et al.*, “Mapping the kinetic Sunyaev-Zel’dovich effect toward MACS J0717.5+3745 with NIKA,” *A& A*, vol. 598, A115, A115, Feb. 2017. arXiv: [1606.07721 \[astro-ph.CO\]](#).
- [219] J. C. Hill, S. Ferraro, N. Battaglia, J. Liu, and D. N. Spergel, “Kinematic Sunyaev-Zel’dovich Effect with Projected Fields: A Novel Probe of the Baryon Distribution with Planck, WMAP, and WISE Data,” *PhRv*, vol. 117, no. 5, 051301, p. 051 301, Jul. 2016. arXiv: [1603.01608 \[astro-ph.CO\]](#).
- [220] S. Ferraro, J. C. Hill, N. Battaglia, J. Liu, and D. N. Spergel, “Kinematic Sunyaev-Zel’dovich effect with projected fields. II. Prospects, challenges, and comparison with simulations,” *PhRvD*, vol. 94, no. 12, 123526, p. 123 526, Dec. 2016. arXiv: [1605.02722 \[astro-ph.CO\]](#).
- [221] E. J. Baxter, B. D. Sherwin, and S. Raghunathan, “Constraining the rotational kinematic Sunyaev-Zel’dovich effect in massive galaxy clusters,” *JCAP*, vol. 2019, no. 6, 001, p. 001, Jun. 2019. arXiv: [1904.04199 \[astro-ph.CO\]](#).
- [222] A. Cooray and X. Chen, “Kinetic Sunyaev-Zeldovich Effect from Halo Rotation,” *ApJ*, vol. 573, no. 1, pp. 43–50, Jul. 2002. arXiv: [astro-ph/0107544 \[astro-ph\]](#).
- [223] J. Chluba and K. Mannheim, “Kinetic Sunyaev-Zeldovich effect from galaxy cluster rotation,” *A& A*, vol. 396, pp. 419–427, Dec. 2002. arXiv: [astro-ph/0208392 \[astro-ph\]](#).
- [224] A. S. Baldi, M. De Petris, F. Sembolini, *et al.*, “Kinetic Sunyaev-Zel’dovich effect in rotating galaxy clusters from MUSIC simulations,” *MNRAS*, vol. 479, no. 3, pp. 4028–4040, Sep. 2018. arXiv: [1805.07142 \[astro-ph.CO\]](#).
- [225] F. De Paolis, V. G. Gurzadyan, A. A. Nucita, *et al.*, “Planck confirmation of the disk and halo rotation of M 31,” *A& A*, vol. 565, L3, p. L3, May 2014. arXiv: [1404.4162](#).
- [226] —, “Triangulum galaxy viewed by Planck,” *A& A*, vol. 593, A57, A57, Sep. 2016. arXiv: [1607.08099](#).

- [227] V. G. Gurzadyan, F. De Paolis, A. A. Nucita, *et al.*, “Planck view of the M 82 galaxy,” *A& A*, vol. 582, A77, A77, Oct. 2015. arXiv: [1509.05212](#).
- [228] ———, “Messier 81’s Planck view versus its halo mapping,” *A& A*, vol. 609, A131, A131, Feb. 2018. arXiv: [1710.04166](#).
- [229] A. H. Maller and J. S. Bullock, “Multiphase galaxy formation: high-velocity clouds and the missing baryon problem,” *MNRAS*, vol. 355, no. 3, pp. 694–712, Dec. 2004. arXiv: [astro-ph/0406632](#) [[astro-ph](#)].
- [230] A. V. Kravtsov, A. A. Vikhlinin, and A. V. Meshcheryakov, “Stellar Mass—Halo Mass Relation and Star Formation Efficiency in High-Mass Halos,” *Astronomy Letters*, vol. 44, no. 1, pp. 8–34, Jan. 2018. arXiv: [1401.7329](#) [[astro-ph.CO](#)].
- [231] R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel, “Concentrations of Dark Halos from Their Assembly Histories,” *ApJ*, vol. 568, no. 1, pp. 52–70, Mar. 2002. arXiv: [astro-ph/0108151](#) [[astro-ph](#)].
- [232] J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel, “Profiles of dark haloes: evolution, scatter and environment,” *MNRAS*, vol. 321, no. 3, pp. 559–575, Mar. 2001. arXiv: [astro-ph/9908159](#) [[astro-ph](#)].
- [233] G. L. Bryan and M. L. Norman, “Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons,” *ApJ*, vol. 495, no. 1, pp. 80–99, Mar. 1998. arXiv: [astro-ph/9710107](#) [[astro-ph](#)].
- [234] J. S. Bullock, A. Dekel, T. S. Kolatt, A. V. Kravtsov, A. A. Klypin, C. Porciani, and J. R. Primack, “A Universal Angular Momentum Profile for Galactic Halos,” *ApJ*, vol. 555, no. 1, pp. 240–257, Jul. 2001. arXiv: [astro-ph/0011001](#) [[astro-ph](#)].
- [235] B. D. Oppenheimer, “Deviations from hydrostatic equilibrium in the circumgalactic medium: spinning hot haloes and accelerating flows,” *MNRAS*, vol. 480, no. 3, pp. 2963–2975, Nov. 2018. arXiv: [1801.00788](#) [[astro-ph.GA](#)].
- [236] S. E. Nuza, F. Parisi, C. Scannapieco, P. Richter, S. Gottlöber, and M. Steinmetz, “The distribution of gas in the Local Group from constrained cosmological simulations: the case for Andromeda and the Milky Way galaxies,” *MNRAS*, vol. 441, no. 3, pp. 2593–2612, Jul. 2014. arXiv: [1403.7528](#) [[astro-ph.GA](#)].
- [237] S. Ferraro and B. Hensley, “Background subtraction uncertainty from submillimetre to millimetre wavelengths,” *MNRAS*, vol. 451, no. 2, pp. 1606–1612, Aug. 2015. arXiv: [1406.2921](#) [[astro-ph.GA](#)].

- [238] M. G. Haehnelt and M. Tegmark, “Using the Kinematic Sunyaev-Zeldovich effect to determine the peculiar velocities of clusters of galaxies.,” *MNRAS*, vol. 279, p. 545, Mar. 1996. arXiv: [astro-ph/9507077](#) [[astro-ph](#)].
- [239] L. Knox, “Determination of inflationary observables by cosmic microwave background anisotropy experiments,” *PhRvD*, vol. 52, no. 8, pp. 4307–4318, Oct. 1995. arXiv: [astro-ph/9504054](#) [[astro-ph](#)].
- [240] N. Sehgal, S. Aiola, Y. Akrami, *et al.*, “CMB-HD: An Ultra-Deep, High-Resolution Millimeter-Wave Survey Over Half the Sky,” *arXiv e-prints*, arXiv:1906.10134, arXiv:1906.10134, Jun. 2019. arXiv: [1906.10134](#) [[astro-ph.CO](#)].
- [241] K. Bundy, M. A. Bershady, D. R. Law, *et al.*, “Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory,” *ApJ*, vol. 798, 7, p. 7, Jan. 2015. arXiv: [1412.1482](#).
- [242] D. A. Wake, K. Bundy, A. M. Diamond-Stanic, *et al.*, “The SDSS-IV MaNGA Sample: Design, Optimization, and Usage Considerations,” *AJ*, vol. 154, 86, p. 86, Sep. 2017. arXiv: [1707.02989](#).
- [243] H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer, “Improving galaxy morphologies for SDSS with Deep Learning,” *MNRAS*, vol. 476, pp. 3661–3676, May 2018. arXiv: [1711.05744](#).
- [244] M. Argudo-Fernández, S. Verley, G. Bergond, *et al.*, “Catalogues of isolated galaxies, isolated pairs, and isolated triplets in the local Universe,” *A&A*, vol. 578, A110, A110, Jun. 2015. arXiv: [1504.00117](#).
- [245] O. Hahn, C. Porciani, C. M. Carollo, and A. Dekel, “Properties of dark matter haloes in clusters, filaments, sheets and voids,” *Monthly Notices of the Royal Astronomical Society*, vol. 375, no. 2, pp. 489–499, Feb. 2007. arXiv: [astro-ph/0610280](#) [[astro-ph](#)].
- [246] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, *et al.*, “Planck 2018 results. I. Overview and the cosmological legacy of Planck,” *arXiv e-prints*, arXiv:1807.06205, arXiv:1807.06205, Jul. 2018. arXiv: [1807.06205](#) [[astro-ph.CO](#)].
- [247] K. R. Stewart, “Gas Accretion and Angular Momentum,” in *Gas Accretion onto Galaxies*, A. Fox and R. Davé, Eds., ser. Astrophysics and Space Science Library, vol. 430, Jan. 2017, p. 249. arXiv: [1612.00513](#) [[astro-ph.GA](#)].
- [248] D. DeFelippis, S. Genel, and G. Bryan, “The Angular Momentum of the Circumgalactic Medium in the Illustris TNG Simulation,” *In prep.*, 2020.
- [249] E. Holmberg, “A photographic photometry of extragalactic nebulae.,” *Meddelanden fran Lunds Astronomi*, vol. 136, p. 1, 1958.

- [250] M. C. Sormani, E. Sobacchi, G. Pezzulli, J. Binney, and R. S. Klessen, “Models of rotating coroneae,” *MNRAS*, vol. 481, no. 3, pp. 3370–3381, Dec. 2018. arXiv: [1809.03437 \[astro-ph.GA\]](#).
- [251] M. A. Strauss and J. A. Willick, “The density and peculiar velocity fields of nearby galaxies,” *Phys. Rep.*, vol. 261, pp. 271–431, Jan. 1995. arXiv: [astro-ph/9502079 \[astro-ph\]](#).
- [252] I. K. Baldry, K. Glazebrook, and S. P. Driver, “On the galaxy stellar mass function, the mass-metallicity relation and the implied baryonic mass function,” *MNRAS*, vol. 388, no. 3, pp. 945–959, Aug. 2008. arXiv: [0804.2892 \[astro-ph\]](#).
- [253] A. R. Duffy, M. J. Meyer, L. Staveley-Smith, M. Bernyk, D. J. Croton, B. S. Koribalski, D. Gerstmann, and S. Westerlund, “Predictions for ASKAP neutral hydrogen surveys,” *MNRAS*, vol. 426, pp. 3385–3402, Nov. 2012. arXiv: [1208.5592](#).
- [254] J. Blue Bird, J. Davis, N. Luber, et al., “CHILES VI: H I and H  $\alpha$  observations for  $z < 0.1$  galaxies; probing H I spin alignment with filaments in the cosmic web,” *MNRAS*, vol. 492, no. 1, pp. 153–176, Feb. 2020.
- [255] M. Birkinshaw and S. F. Gull, “A test for transverse motions of clusters of galaxies,” *Nature*, vol. 302, no. 5906, pp. 315–317, Mar. 1983.
- [256] S. C. Hotinli, J. Meyers, N. Dalal, A. H. Jaffe, M. C. Johnson, J. B. Mertens, M. Münchmeyer, K. M. Smith, and A. van Engelen, “Transverse Velocities with the Moving Lens Effect,” *PhRv*, vol. 123, no. 6, 061301, p. 061 301, Aug. 2019. arXiv: [1812.03167 \[astro-ph.CO\]](#).
- [257] S. Yasini, N. Mirzatury, and E. Pierpaoli, “Pairwise Transverse Velocity Measurement with the Rees-Sciama Effect,” *ApJL*, vol. 873, no. 2, L23, p. L23, Mar. 2019. arXiv: [1812.04241 \[astro-ph.CO\]](#).
- [258] R. Hagala, C. Llinares, and D. F. Mota, “The slingshot effect as a probe of transverse motions of galaxies,” *A&A*, vol. 628, A30, A30, Aug. 2019. arXiv: [1907.01429 \[astro-ph.GA\]](#).
- [259] E. Tempel and N. I. Libeskind, “Galaxy Spin Alignment in Filaments and Sheets: Observational Evidence,” *ApJL*, vol. 775, no. 2, L42, p. L42, Oct. 2013. arXiv: [1308.2816 \[astro-ph.CO\]](#).
- [260] A. Krolewski, S. Ho, Y.-C. Chen, P. F. Chan, A. Tenneti, D. Bizyaev, and K. Kraljic, “Alignment between Filaments and Galaxy Spins from the MaNGA Integral-field Survey,” *ApJ*, vol. 876, no. 1, 52, p. 52, May 2019. arXiv: [1902.09797 \[astro-ph.GA\]](#).
- [261] C. Welker, J. Bland-Hawthorn, J. Van de Sande, et al., “The SAMI Galaxy Survey: first detection of a transition in spin orientation with respect to cosmic filaments in the stellar



- kinematics of galaxies,” *MNRAS*, vol. 491, no. 2, pp. 2864–2884, Jan. 2020. arXiv: [1909.12371 \[astro-ph.GA\]](#).
- [262] B. Ménard, R. Scranton, M. Fukugita, and G. Richards, “Measuring the galaxy-mass and galaxy-dust correlations through magnification and reddening,” *MNRAS*, vol. 405, no. 2, pp. 1025–1039, Jun. 2010. arXiv: [0902.4240 \[astro-ph.CO\]](#).
- [263] J. B. Melin, J. G. Bartlett, Z. Y. Cai, et al., “Dust in galaxy clusters: Modeling at millimeter wavelengths and impact on Planck cluster cosmology,” *A&A*, vol. 617, A75, A75, Sep. 2018. arXiv: [1808.06807 \[astro-ph.CO\]](#).
- [264] A. Amekhyan, “On the structure of galactic halos and the microwave temperature maps,” *arXiv e-prints*, arXiv:1903.10757, arXiv:1903.10757, Mar. 2019. arXiv: [1903.10757 \[astro-ph.GA\]](#).
- [265] W. Hu, “Dark energy and matter evolution from lensing tomography,” *PhRvD*, vol. 66, no. 8, 083515, p. 083 515, Oct. 2002. arXiv: [astro-ph/0208093 \[astro-ph\]](#).
- [266] M. C. Neyrinck, I. Szapudi, and A. S. Szalay, “Rejuvenating the Matter Power Spectrum: Restoring Information with a Logarithmic Density Mapping,” *ApJL*, vol. 698, no. 2, pp. L90–L93, Jun. 2009. arXiv: [0903.4693 \[astro-ph.CO\]](#).
- [267] H.-J. Seo, M. Sato, S. Dodelson, B. Jain, and M. Takada, “Re-capturing Cosmic Information,” *ApJL*, vol. 729, no. 1, L11, p. L11, Mar. 2011. arXiv: [1008.0349 \[astro-ph.CO\]](#).
- [268] M. Shirasaki, “Statistical connection of peak counts to power spectrum and moments in weak-lensing field,” *MNRAS*, vol. 465, no. 2, pp. 1974–1983, Feb. 2017. arXiv: [1610.00840 \[astro-ph.CO\]](#).
- [269] M. Takada, “Statistical challenges in weak lensing cosmology,” in *Statistical Challenges in 21st Century Cosmology*, A. Heavens, J.-L. Starck, and A. Krone-Martins, Eds., ser. IAU Symposium, vol. 306, May 2014, pp. 78–89. arXiv: [1407.3330 \[astro-ph.CO\]](#).
- [270] H. W. Lin, M. Tegmark, and D. Rolnick, “Why Does Deep and Cheap Learning Work So Well?” *Journal of Statistical Physics*, vol. 168, no. 6, pp. 1223–1247, Sep. 2017. arXiv: [1608.08225 \[cond-mat.dis-nn\]](#).
- [271] A. S. Miller, “A review of neural network applications in Astronomy,” *Vistas in Astronomy*, vol. 36, no. 2, pp. 141–161, Jan. 1993.
- [272] R. Tagliaferri, G. Longo, L. Milano, et al., “Neural networks in astronomy,” *Neural Networks*, vol. 16, no. 3, pp. 297–319, 2003, Neural Network Analysis of Complex Scientific Data: Astronomy and Geosciences.

- [273] S. Ravanbakhsh, J. Oliva, S. Fromenteau, et al., “Estimating cosmological parameters from the dark matter distribution,” in Proceedings of The 33rd International Conference on Machine Learning, M. F. Balcan and K. Q. Weinberger, Eds., ser. Proceedings of Machine Learning Research, vol. 48, PMLR, Jun. 2016, pp. 2407–2416.
- [274] B. Jain and U. Seljak, “Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes,” ApJ, vol. 484, no. 2, pp. 560–573, Jul. 1997. arXiv: [astro-ph/9611077](#) [[astro-ph](#)].
- [275] W. Fang and Z. Haiman, “Constraining dark energy by combining cluster counts and shear-shear correlations in a weak lensing survey,” PhRvD, vol. 75, no. 4, 043010, p. 043 010, Feb. 2007. arXiv: [astro-ph/0612187](#) [[astro-ph](#)].
- [276] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” arXiv e-prints, arXiv:1207.0580, arXiv:1207.0580, Jul. 2012. arXiv: [1207.0580](#) [[cs.NE](#)].
- [277] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv e-prints, arXiv:1412.6980, arXiv:1412.6980, Dec. 2014. arXiv: [1412.6980](#) [[cs.LG](#)].
- [278] M. Kilbinger, K. Benabed, J. Guy, P. Astier, et al., “Dark-energy constraints and correlations with systematics from CFHTLS weak lensing, SNLS supernovae Ia and WMAP5,” A&A, vol. 497, no. 3, pp. 677–688, Apr. 2009. arXiv: [0810.5129](#) [[astro-ph](#)].
- [279] P. Alfeld and other, “A trivariate clough-tocher scheme for tetrahedral data,” Defense Technical Information vol. Kingma, D. P. and Ba, J. 1984.
- [280] R. Renka and A. Cline, “A triangle-based c1 interpolation method,” Rocky Mountain J. Math, vol. 14, no. 1, 1984.
- [281] J. A. Nelder and R. Mead, “A simplex method for function minimization,” The computer journal, vol. 7, no. 4, pp. 308–313, 1965.
- [282] M. Sato, K. Ichiki, and T. T. Takeuchi, “Precise Estimation of Cosmological Parameters Using a More Accurate Likelihood Function,” PhRv, vol. 105, no. 25, 251301, p. 251 301, Dec. 2010. arXiv: [1011.4996](#) [[astro-ph.CO](#)].
- [283] C. Chang, M. Jarvis, B. Jain, et al., “The effective number density of galaxies for weak lensing measurements in the LSST project,” MNRAS, vol. 434, no. 3, pp. 2121–2135, Sep. 2013. arXiv: [1305.0793](#) [[astro-ph.CO](#)].
- [284] E. Semboloni, H. Hoekstra, and J. Schaye, “Effect of baryonic feedback on two- and three-point shear statistics: prospects for detection and improved modelling,” MNRAS, vol. 434, no. 1, pp. 148–162, Sep. 2013. arXiv: [1210.7303](#) [[astro-ph.CO](#)].



- [285] K. Osato, M. Shirasaki, and N. Yoshida, “Impact of Baryonic Processes on Weak-lensing Cosmology: Power Spectrum, Nonlocal Statistics, and Parameter Bias,” *ApJ*, vol. 806, no. 2, p. 186, Jun. 2015. arXiv: [1501.02055 \[astro-ph.CO\]](#).
- [286] H. Hildebrandt, M. Viola, C. Heymans, et al., “KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing,” *MNRAS*, vol. 465, no. 2, pp. 1454–1498, Feb. 2017. arXiv: [1606.05338 \[astro-ph.CO\]](#).
- [287] T. M. C. Abbott, F. B. Abdalla, S. Avila, et al., “Dark Energy Survey year 1 results: Constraints on extended cosmological models from galaxy clustering and weak lensing,” *PhRvD*, vol. 99, no. 12, 123505, p. 123 505, Jun. 2019. arXiv: [1810.02499 \[astro-ph.CO\]](#).
- [288] T. Lu and Z. Haiman, “The matter fluctuation amplitude inferred from the weak lensing power spectrum and correlation function in CFHTLenS data,” *MNRAS*, vol. 490, no. 4, pp. 5033–5042, Dec. 2019. arXiv: [1905.07803 \[astro-ph.CO\]](#).
- [289] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps,” *arXiv e-prints*, arXiv:1312.6034, arXiv:1312.6034, Dec. 2013. arXiv: [1312.6034 \[cs.CV\]](#).
- [290] G. Montavon, S. Bach, A. Binder, W. Samek, and K.-R. Müller, “Explaining NonLinear Classification Decisions with Deep Taylor Decomposition,” *arXiv e-prints*, arXiv:1512.02479, arXiv:1512.02479, Dec. 2015. arXiv: [1512.02479 \[cs.LG\]](#).
- [291] A. Shrikumar, P. Greenside, A. Shcherbina, and A. Kundaje, “Not Just a Black Box: Learning Important Features Through Propagating Activation Differences,” *arXiv e-prints*, arXiv:1605.01713, arXiv:1605.01713, May 2016. arXiv: [1605.01713 \[cs.LG\]](#).
- [292] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, “SmoothGrad: removing noise by adding noise,” *arXiv e-prints*, arXiv:1706.03825, arXiv:1706.03825, Jun. 2017. arXiv: [1706.03825 \[cs.LG\]](#).
- [293] P. Virtanen, R. Gommers, T. E. Oliphant, et al., “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [294] H. J. Kelley, “Gradient theory of optimal flight paths,” *Ars Journal*, vol. 30, no. 10, pp. 947–954, 1960.
- [295] M. Sundararajan, A. Taly, and Q. Yan, “Axiomatic Attribution for Deep Networks,” *arXiv e-prints*, arXiv:1703.01365, arXiv:1703.01365, Mar. 2017. arXiv: [1703.01365 \[cs.LG\]](#).
- [296] M. D. Zeiler, “Adaptive deconvolutional networks for mid and high level feature learning,” Nov. 2011, pp. 2018–2025.

- [297] M. Alber, S. Lapuschkin, P. Seegerer, et al., “iNNvestigate neural networks!” arXiv e-prints, arXiv:1808.04260, arXiv:1808.04260, Aug. 2018. arXiv: [1808.04260 \[cs.LG\]](#).
- [298] B. Kim, J. Seo, S. Jeon, et al., “Why are Saliency Maps Noisy? Cause of and Solution to Noisy Saliency Maps,” arXiv e-prints, arXiv:1902.04893, arXiv:1902.04893, Feb. 2019. arXiv: [1902.04893 \[cs.LG\]](#).
- [299] A. Ghorbani, A. Abid, and J. Zou, “Interpretation of Neural Networks is Fragile,” arXiv e-prints, arXiv:1710.10547, arXiv:1710.10547, Oct. 2017. arXiv: [1710.10547 \[stat.ML\]](#).
- [300] P.-J. Kindermans, S. Hooker, J. Adebayo, et al., “The (Un)reliability of saliency methods,” arXiv e-prints, arXiv:1711.00867, arXiv:1711.00867, Nov. 2017. arXiv: [1711.00867 \[stat.ML\]](#).
- [301] W. Nie, Y. Zhang, and A. Patel, “A Theoretical Explanation for Perplexing Behaviors of Backpropagation-based Visualizations,” arXiv e-prints, arXiv:1805.07039, arXiv:1805.07039, May 2018. arXiv: [1805.07039 \[cs.CV\]](#).
- [302] J. Herault and C. Jutten, “Space or time adaptive signal processing by neural network models,” in American Institute of Physics Conference Series, ser. American Institute of Physics Conference Series, vol. 151, Aug. 1986, pp. 206–211.
- [303] P. Comon, “Independent component analysis, a new concept?” Signal Processing, vol. 36, no. 3, pp. 287–314, Apr. 1994.
- [304] G. Fasano and A. Franceschini, “A multidimensional version of the Kolmogorov-Smirnov test,” MNRAS, vol. 225, pp. 155–170, Mar. 1987.

## Appendix A: Gaussian likelihood approximation

One way to assess how valid the Gaussian approximation is for the likelihood of a given observable is to estimate its probability density function (PDF) from our simulations without assuming any specific functional form. A non-parametric method to do that estimation is the Kernel Density Estimator (KDE). The main challenge to apply this approach to lensing peaks is how to achieve a density estimator in a high dimensional space with a limited number of independent vectors (512 per model).

We performed an analysis with noisy data within the framework of a different study, that supports that a Gaussian likelihood is not a bad approximation for lensing peaks. The dataset corresponds to the same cosmologies used for this study, and the convergence maps have been smoothed with a characteristic scale of 1 arcmin, but they also have an ellipticity noise of  $\sigma_\epsilon = 0.4$  present. To reduce the dimensionality of the observable, we performed an Independent Component Analysis (ICA) [302, 303]. This method provides the directions that maximize negative entropy, which can be interpreted as the directions in which the data is less Gaussian. As a pre-processing step, we whitened the data (i.e. we removed its mean and normalized its covariance), and then we projected the whitened data into the 9 directions found following ICA. We then used a KDE to estimate the PDF of the resulting data. While we found some non-Gaussianities, specially for peak counts corresponding to high significance, the effect on the likelihood (and corresponding credible contours) is limited.

As an illustration, in Fig. A.1 we show the difference in credible contours obtained from a Gaussian likelihood from those obtained using a KDE. We display only the contours derived using only peaks with a signal-to-noise greater than 3. These are the peaks for which the non-Gaussianities are the most pronounced, and yet the contours obtained with both methods are comparable. Using a model to predict peak counts that does not rely on N-body simulations, [170] also

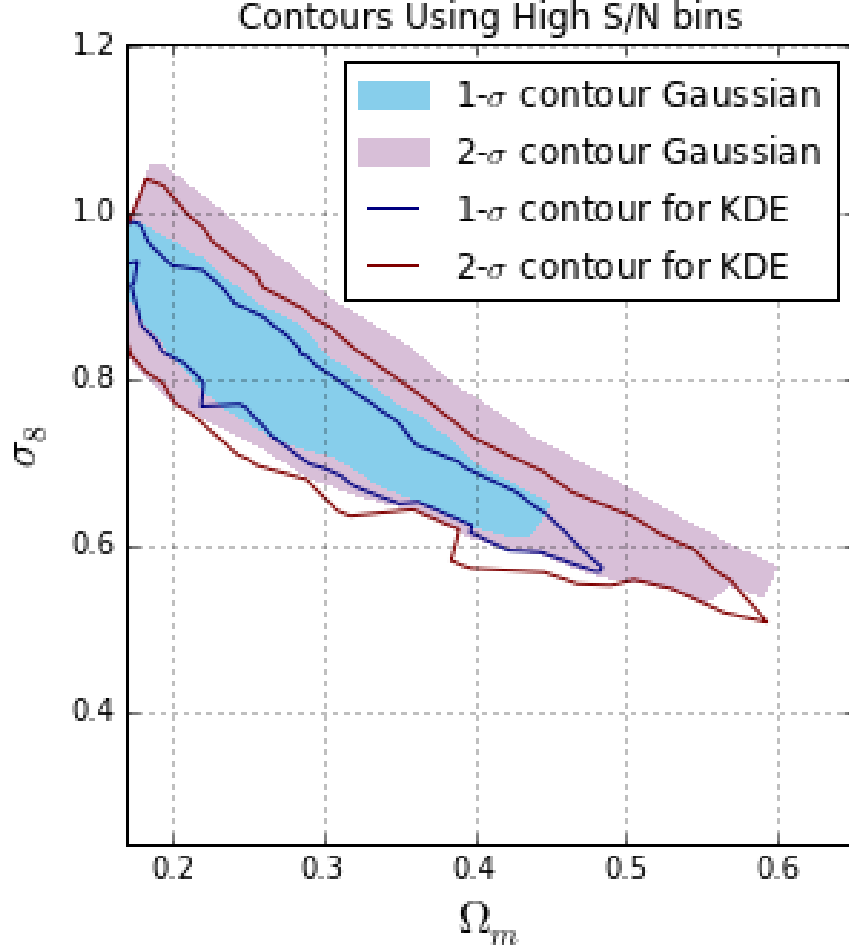


Figure A.1: *Credible contours for  $\{\Omega_m, \sigma_8, \Sigma_8\}$  from lensing peak counts on noisy  $\kappa$  maps. Filled contours correspond to a Gaussian likelihood, and solid lines to contours corresponding to KDE estimates.*

found that a Gaussian likelihood is a good approximation (to  $\sim 10\%$ ) for lensing peaks.

To analyze whether a Gaussian distribution is a good approximation for the  $\{\Omega_m, \sigma_8\}$  predictions from the neural network we used a modification of the Kolmogorov-Smirnoff test that can be applied to two-dimensional distributions [304]. For each model, we computed the mean and covariance from the predictions for the test maps. Then, we tested the predictions against a Gaussian distribution defined by the estimated mean and covariance.

The null hypothesis, that there is no statistical difference between the distribution of our empirical samples (neural network predictions) and a Gaussian, cannot be rejected with a confidence

of 99% except for 2 models which are far from the fiducial,  $\{\Omega_m = 0.450, \sigma_8 = 0.200\}$  and  $\{\Omega_m = 0.452, \sigma_8 = 0.454\}$ . We conclude that a Gaussian likelihood is a reasonable approximation for the predictions from the neural network.

## Appendix B: Sensitivity of results to interpolation

To assess how sensitive our results were to the models sampled from the parameter space  $\{\Omega_m, \sigma_8\}$ , we trained an additional network on the same un-smoothed  $\kappa$  maps but removing the model  $\{\Omega_m = 0.261, \sigma_8 = 0.802\}$  from the training data set. When fed the test maps for that cosmology, the network that was not exposed to it during training yielded somewhat different predictions than the network which had seen maps from that model during training. The differences in the mean prediction were very small, with a shift of  $-1.0\%$  in  $\Omega_m$  and  $-0.1\%$  in  $\sigma_8$ . The change in scatter is more significant, the standard deviation in the predictions for  $\Omega_m$  increasing by  $80.8\%$  and that for the  $\sigma_8$  predictions by  $12.2\%$ . The larger degradation for  $\Omega_m$  may be related with the fact that the network’s architecture seems to have greater difficulty in distinguishing between models that differ in that parameter, as was shown in § 6.4 for both GRFs and smoothed convergence maps.

While this sensitivity to interpolation highlights how relevant a well-sampled training data set is for proper generalization by the network’s architecture, we are mostly concerned about how interpolation errors propagate into the inferred parameters’ constraints. That effect is small, as Fig. B.1 shows. The credible contours inferred from the predictions by both networks barely change, and the same applies to the marginal distributions inferred for both  $\Omega_m$  and  $\sigma_8$ . We show the contours computed for the worst-case scenario, that is, when the model missing from the training data-set is the “true” cosmology.

The small change in the parameter constraints’ from both networks indicate that our main conclusions would not change with a different sampling of the parameter space. Besides, as the priors on our cosmological parameters improve with new experiments, the parameter volume to be explored will shrink and the number of models that need to be simulated to sample that space properly will also decrease.

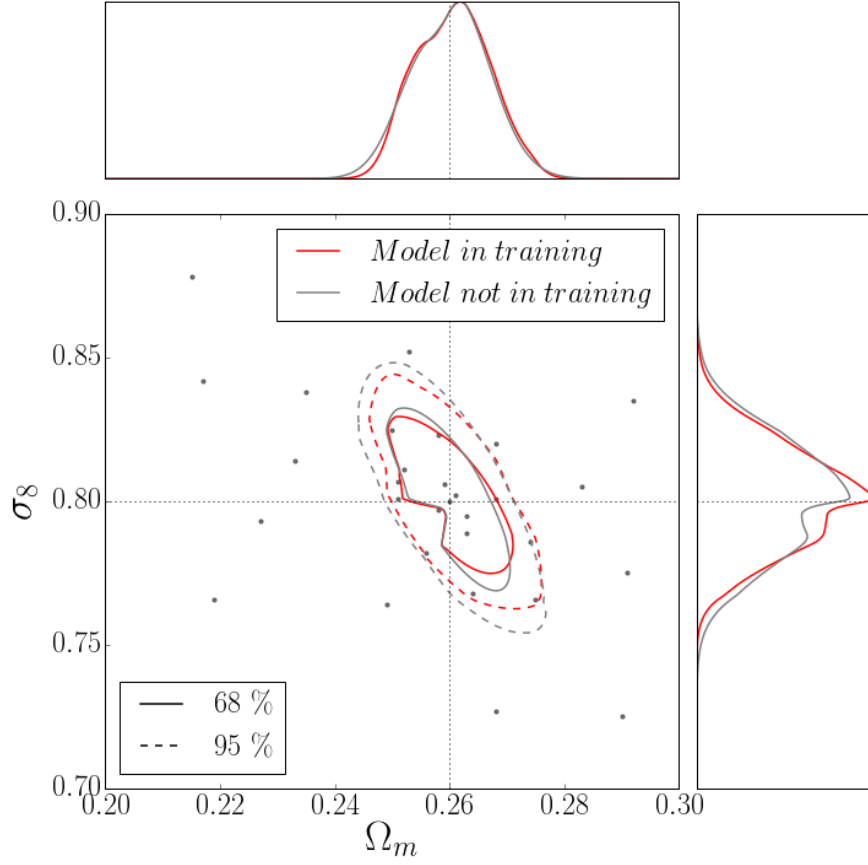


Figure B.1: 68% and 95% credible contours for un-smoothed ( $\approx 0.2$  arcmin/pixel)  $\kappa$  maps, derived from two neural networks with the same architecture: the original one trained on all 96 cosmologies (red) and another one for which the model  $\{\Omega_m = 0.261, \sigma_8 = 0.802\}$  was excluded (grey). The assumed true value ( $\{\Omega_m = 0.261, \sigma_8 = 0.802\}$ ) is indicated by black dotted lines. The upper and right panels show the marginal distribution for  $\Omega_m$  and  $\sigma_8$ , respectively.

## Appendix C: Impact of filter misalignment and centering errors.

Even in the absence of noise, the true temperature dipole induced by the rkSZ effect can differ from the one measured for any given galaxy. A centering error in the aperture filter described in § 5.3.1, and/or a misalignment between the filter’s axis and the galaxy’s projected spin vector, will suppress the measured dipole. The same applies when convolving a matched filter with the CMB data. While for a single galaxy the maximum response to the matched filter localizes the center of the galaxy’s halo, when stacking the data for many noise-dominated galaxies, their center needs to be chosen a priori.

To assess the sensitivity of these filters to these errors, we computed the mean response of the filter according to Eq. 5.19, assuming that the centering offsets and misalignments both follow zero-mean, normal distributions:

$$\langle s \rangle_\theta = \frac{2}{\sqrt{2\pi}\sigma_\theta} \int_0^\infty d\theta \exp\left[-\frac{\theta^2}{2\sigma_\theta^2}\right] s(\theta), \quad (\text{C.1})$$

$$\langle s \rangle_{xy} = \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{dx dy}{2\pi\sigma_{xy}^2} \exp\left[-\frac{x^2 + y^2}{2\sigma_{xy}^2}\right] s(x, y). \quad (\text{C.2})$$

We show in Fig. C.1 the mean effect on the measured signal (dipole for the aperture filter and maximum correlation for matched filter) of a filter decentering and misalignment for three galaxies of different mass, the two different atmosphere models described in S 5.2, and three different CMB experiment configurations.

The sensitivity of both filters to errors in the galaxies’ spin angle estimation is similar, and does not depend strongly on the beam resolution of the CMB experiment (see lower panels of Fig. C.1). Even with misalignment errors with FWHM=90 deg, the signal measured by the filters will be suppressed by less than 30% relative to its true value, regardless of the galaxies’ mass, the CMB



beam resolution and the galactic atmosphere model used to predict the signal.

The sensitivity to a filter decentering is strongly dependent on the resolution of the CMB experiments, the matched filter being slightly more robust than the aperture filter, in particular for high-resolution CMB experiments such as CMB-S4 (1 arcmin beam). Still, for the worse case scenario, which corresponds to a high resolution experiment using an aperture filter and galaxy atmospheres that follow a hot fast rotator model, a decentering error with a FWHM= $0.2 R_{\text{vir}}$  (which corresponds to  $\sim 1$  arcmin for a MaNGA-like survey, or the halo center falling outside of the galaxy) suppresses the measured signal by less than 40%.

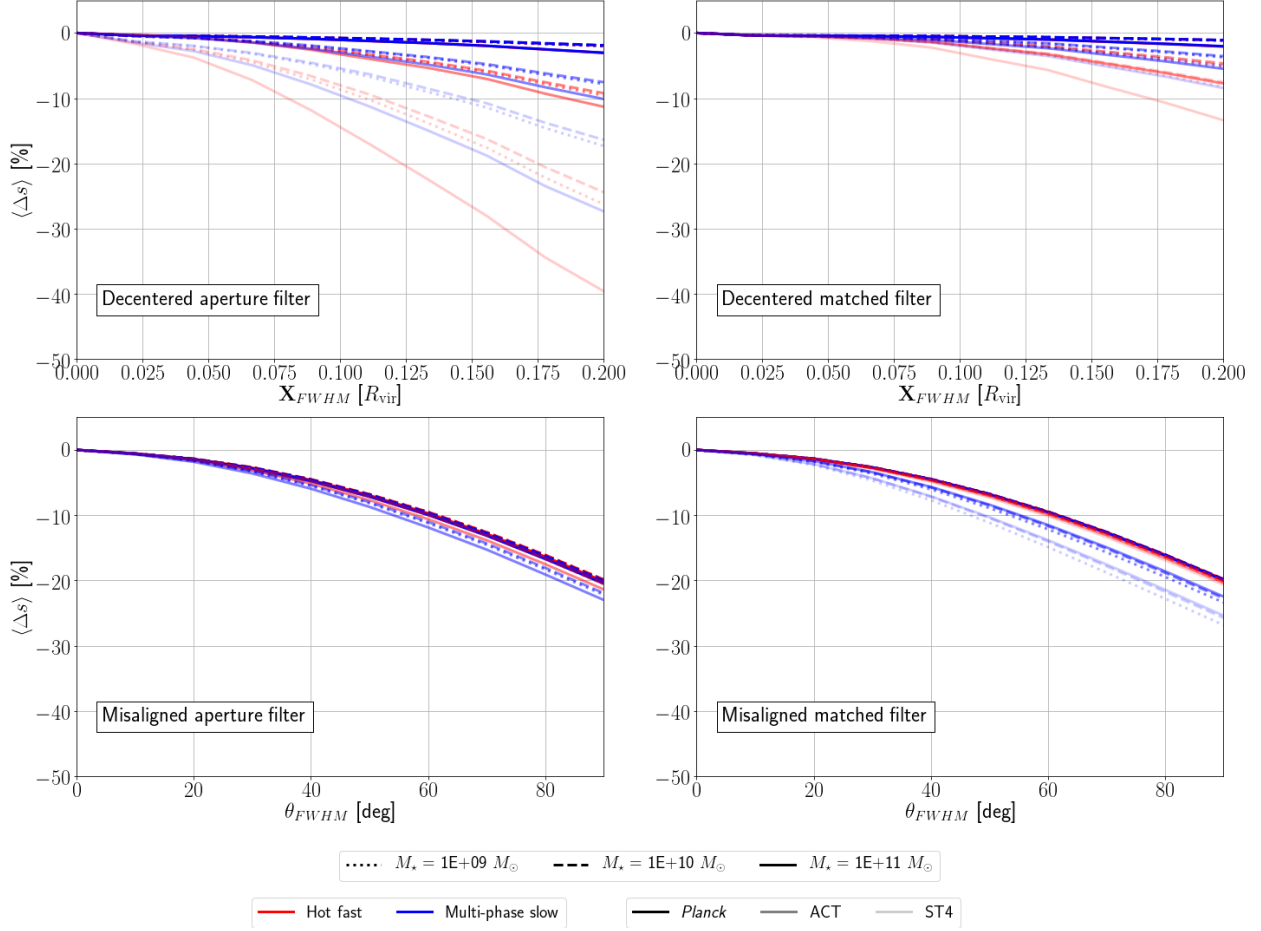


Figure C.1: Mean suppression in the measured rkSZ signal due to decentering relative to the halo position and to misalignment relative to the galaxy’s projected spin angle. Color indicates whether the rkSZ signal corresponds to that of a hot, fast or cold, slow rotator. The intensity of the color indicates a different CMB experiment configuration (*Planck* for strong color, CMB-S4 for the faintest color and ACT for the intermediate intensity). Finally, the type of line used indicates the mass of the galaxy. **Upper panels:** Effect of a Gaussian error in the position of the filter, relative to the halo’s center, as a function of the error’s FWHM in units of the virial radius. On the left, effect for an aperture filter measuring the signal’s dipole. On the right, effect for the measured correlation of a matched filter and the signal at the estimated (erroneous) halo center. **Lower panels:** Effect of a Gaussian error in the orientation of the filter axis relative to the galaxy’s projected spin vector, as a function of the error’s FWHM in degrees. As in the upper panels, on the left the effect for the aperture filter is displayed and on the right, for the matched filter.

## Appendix D: Variance of aperture filter dipole measurements

Following [237], we can estimate the variance of a dipole measured over an aperture. For any given galaxy, the dipole measurement is  $s = \overline{\Delta T}_R - \overline{\Delta T}_L$ , where  $\overline{\Delta T}_R$  is the mean temperature anisotropy measured within the right half of the aperture and  $\overline{\Delta T}_L$  the same within the left half of the aperture. The CMB anisotropies have rotational symmetry, and the variance on the measured signal induced by them is given by Eqs. 5.13 and 5.14.

The window function used in this study is a semi-circle centered on each galaxy of radius  $a$  in units of the its host halo's virial radius projected on the sky. This choice of window function is not circularly symmetric and it can be thought of the product of a top hat and a rectangular filter,  $W(x, y) = W_1(x, y)W_2(x, y)$ ,

$$W_1(x, y) = \begin{cases} \frac{1}{\pi a^2} & \text{if } x^2 + y^2 \leq a^2 \\ 0 & \text{otherwise} \end{cases} \quad (\text{D.1})$$

$$W_2(x, y) = \begin{cases} 1 & \text{if } |y| \leq a \wedge x \leq a \wedge x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{D.2})$$

for a right aperture. For a left aperture,  $W_2$  is displaced by  $a$  to the left of the  $x$ -axis. The resulting half circle's window form, as a function of  $\ell \equiv (\ell_x, \ell_y)$  is  $\widetilde{W} = \widetilde{W}_1 * \widetilde{W}_2$ , or

$$\widetilde{W}(\ell) = \frac{8}{a\ell} J_1(a\ell) * \frac{\sin\left(\frac{a\ell_x}{2}\right) \sin(a\ell_y)}{\ell_x \ell_y} \exp\left[\mp i \frac{a\ell_x}{2}\right], \quad (\text{D.3})$$

here  $J_1$  is the Bessel function of first kind,  $*$  represents a convolution, the negative sign on the exponential corresponds to the right aperture and the positive sign to the left aperture. We assume

a Gaussian beam function which depends on the CMB experiment's beam's full width at half maximum (FWHM),  $b_\ell = \exp \left[ -\frac{FWHM^2}{16 \ln 2} \ell(\ell + 1) \right]$ . The power spectrum  $C_\ell$  includes that of the CMB and any contributions of instrumental noise.

## Appendix E: Selecting galaxies from MaNGA for stacking.

We selected a set of galaxies for stacking by combining information from the MaNGA DRPALL catalog with two value-added catalogs from SDSS DR15: the MaNGA Morphology Deep Learning DR15 and the GEMA-VAC, see § 5.5 for a brief description of them. The starting point are the 4,690 records with information in DRPALL. Removing all records flagged with potential quality issues reduces the initial number to 4,196 (see <https://www.sdss.org/dr15/algorithms/bitmasks/> for a description of the bitmask used in the `drp3qual` field).

We removed objects with more than one observation, that is, duplicates in the `manga_id` field. There are 4,093 un-flagged objects with unique observations. Only objects in one of the three science target samples were considered, bringing the total number to 3,939. We computed their projected spin angle as described in § 5.5.3, and removed those galaxies for which such a calculation yielded numerical errors, keeping 3,931 galaxies.

A key assumption in the analysis of *Planck* data on MaNGA galaxies is that the spin of the outer gaseous halo is aligned with that of the inner regions, which are the ones probed by MaNGA IFU spectrographs. We deemed this assumption more likely if the inner kinematics probed by different tracers are consistent with each other. To test for consistency, we also computed the spin angle using the  $H_\alpha$  line (6,564 Å). We modeled the difference between this angle and the one from the O II line by a random variable whose pdf is a combination of a (zero mean) Gaussian and a uniform distributions, as is shown on Fig. E.1. The rationale for this choice is that, while most galaxies show a high correlation between spins estimated with different emission lines, some show little or no correlation. We calculated the best fit values for the Gaussian width and the uniform distribution height, and used the former to discard galaxies for which the difference between the spin angle estimated from O II and that from  $H_\alpha$  exceeds five standard deviations of the Gaussian component. The standard deviation that maximizes the likelihood of our data is 1.8 deg (although

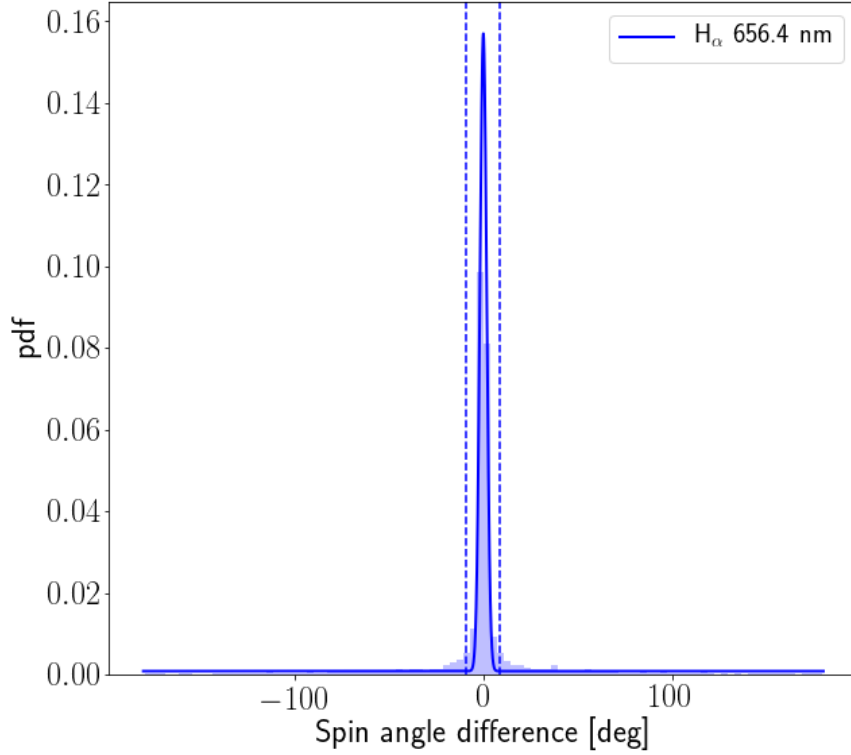


Figure E.1: Histogram showing the difference between the spin angle computed using the  $H_\alpha$  line and the angle derived from O II (average of both O II lines). A mixture model with a Gaussian and a uniform component is a poor fit (indicated by the fat tails). To compensate for the badness of fit, we apply a  $5\sigma$  cut to identify outliers.

the fit is not good, see Fig. E.1), which indicates a very tight correlation between the spin angle measured using  $H_\alpha$  and O II. If that standard deviation is representative of the true uncertainty on the spin angle, our measurements will not be severely affected by errors in the spin angle estimation (see Appendix C). A visual inspection of some of the outliers showed that they were either galaxies with a complex velocity field (i.e. no clear overall rotation pattern) or face-on systems that would contribute little to a rkSZ measurement. Removing the outliers shrank our stacking sample to 2,901 galaxies.

Adding morphological information allowed us to reject galaxies with a high probability of being interactive systems ( $P_{\text{MERG}} > 0.95$ ). This was motivated by the fact that the kinematics of the outer regions of interacting systems can be perturbed to the point of having little correlation with the inner kinematics probed by MaNGA. This further reduced the size of our stacking sample

to a final number of 2,664 galaxies.