

ANALYSIS METHODS FOR PRECISION COSMOLOGY WITH WEAK GRAVITATIONAL LENSING

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Science and Engineering

2022

Robin E. Upham

Jodrell Bank Centre for Astrophysics
Department of Physics and Astronomy
School of Natural Sciences

Table of contents

Abstract	9
Declaration	11
Copyright statement	13
Acknowledgements	15
Supporting publications	17
Supporting code and data	19
1. Cosmology	21
1.1. Standard cosmological model	21
1.1.1. Cosmological principle and anisotropies	21
1.1.2. Cosmic expansion and redshift	23
1.1.3. The FLRW Universe	24
1.1.4. Constituents of the Universe	27
1.1.4.1. Baryonic matter	27
1.1.4.2. Dark matter	27
1.1.4.3. Dark energy	29
1.1.4.3.1. Evidence for dark energy	29
1.1.4.3.2. Models of dark energy	29
1.1.4.4. Other constituents	31
1.1.5. Expansion history	31
1.1.5.1. Inflation	31
1.1.5.2. Radiation domination	32
1.1.5.3. Matter domination	32
1.1.5.4. Dark energy domination	33
1.1.6. Density parameters	33
1.1.7. Free parameters	34
1.2. Outstanding questions	35
1.3. Observational probes	37
1.3.1. Cosmic microwave background	37
1.3.2. Gravitational lensing	40
1.3.3. Type Ia supernovae	41
1.3.4. Baryon acoustic oscillations	41
1.3.5. Redshift-space distortions	42
1.3.6. Gravitational waves	42
1.3.7. Hydrogen 21 cm line	43
1.4. Weak gravitational lensing	43
1.4.1. Combination with galaxy clustering	44

1.4.2.	Challenges in weak lensing	45
1.4.3.	Observational history of weak lensing	47
1.4.4.	<i>Euclid</i> satellite	49
1.4.5.	Other future surveys	51
2.	Cosmological estimators and likelihoods	55
2.1.	Cosmological fields on the sphere	55
2.1.1.	Spin	56
2.2.	Weak lensing fields and two-point statistics	58
2.2.1.	Shear field	58
2.2.2.	Galaxy number overdensity field	59
2.2.3.	3×2 pt power spectra	60
2.2.4.	Shape and shot noise	61
2.3.	Pseudo- C_ℓ method	62
2.4.	Bayesian inference and likelihoods	63
2.4.1.	Credible regions and sigma notation	64
2.4.2.	Gaussian likelihood and covariance	65
3.	Exact likelihood of pseudo-C_ℓ estimates from Gaussian fields	67
3.1.	Introduction	67
3.2.	Pseudo- $a_{\ell m}$ distribution	68
3.2.1.	Full-sky $a_{\ell m}$ distribution	68
3.2.2.	Effect of a cut sky	69
3.2.3.	Pseudo- $a_{\ell m}$ distribution	70
3.3.	Pseudo- C_ℓ distribution	73
3.3.1.	The pseudo- C_ℓ estimator	74
3.3.2.	The joint distribution of quadratic forms	76
3.3.2.1.	The likelihood of a subset of pseudo- C_ℓ estimates	77
3.3.2.2.	The likelihood of Quadratic Maximum Likelihood estimates	77
3.4.	Application to cosmic microwave background polarisation	78
3.4.1.	CMB pseudo- $a_{\ell m}$ covariance	79
3.4.2.	Implementation	80
3.4.3.	Simulations	81
3.4.4.	Comparison to approximation	81
3.5.	Results	82
3.5.1.	Marginal distributions	82
3.5.2.	Correlation between spectra	84
3.5.2.1.	Comparison to approximation	84
3.6.	Conclusions	90
4.	Sufficiency of a Gaussian likelihood for weak lensing power spectra	93
4.1.	Introduction	93

4.2.	Full-sky likelihood	94
4.2.1.	Background	94
4.2.1.1.	Wishart distribution	94
4.2.1.2.	Gaussian distribution	95
4.2.2.	Full sky: Methodology	96
4.2.2.1.	Theory	97
4.2.2.2.	Mock observations	97
4.2.2.3.	Likelihoods	98
4.2.3.	Full sky: Summary statistics	98
4.2.3.1.	Posterior maximum	99
4.2.3.2.	Posterior mean and standard deviation	99
4.2.4.	Full sky: Posterior contours	100
4.2.4.1.	Baseline setup	101
4.2.4.2.	Robustness to deviation from baseline setup	101
4.3.	Cut-sky likelihood	104
4.3.1.	Cut sky: Effect on marginal distributions	105
4.3.1.1.	Impact of additional non-Gaussianity	107
4.3.2.	Cut sky: Effect on dependence structure	109
4.3.2.1.	Mutual information	110
4.3.2.2.	Impact of additional non-Gaussian dependence	113
4.3.2.3.	Transcovariance	114
4.4.	Non-Gaussian fields	115
4.4.1.	Non-Gaussian fields: Simulations	117
4.4.2.	Non-Gaussian fields: Effect on marginal distributions	117
4.4.3.	Non-Gaussian fields: Effect on dependence structure	118
4.5.	Conclusions	119
5.	Covariance of weak lensing pseudo-C_ℓ estimates	121
5.1.	Introduction	121
5.2.	Cosmic shear power spectrum covariance contributions	122
5.2.1.	Gaussian covariance	123
5.2.2.	Super-sample covariance	125
5.2.3.	Connected non-Gaussian covariance	126
5.3.	Comparison to simulations	127
5.3.1.	Method	127
5.3.1.1.	Choice of redshift distribution	128
5.3.2.	Results	129
5.4.	Importance of covariance components and dependence on mask	133
5.4.1.	Relative sizes of components	133
5.4.1.1.	Without shape noise	133
5.4.1.2.	With shape noise	133

5.4.2.	Importance for parameter constraints	135
5.4.2.1.	Connected non-Gaussian approximation	136
5.4.2.2.	Results	137
5.4.2.3.	Effect of marginalisation over additional parameters	140
5.5.	Conclusions	142
6.	Dependence of cosmological parameter constraints on angular binning of weak lensing two-point statistics	145
6.1.	Introduction	145
6.2.	Dependence of posterior uncertainties on number of angular bins	146
6.2.1.	Measurement of posterior uncertainties	146
6.2.1.1.	Modelling of two-point statistics	146
6.2.1.2.	Likelihood and covariance	148
6.2.1.3.	Cut-sky treatment	149
6.2.2.	Posterior uncertainty as a function of number of angular bins	150
6.3.	Exploration of additional effects	152
6.3.1.	Impact of f_{sky} approximation	154
6.3.2.	Contrast between power spectrum and correlation function behaviour for small numbers of bins	155
6.3.3.	Impact of noise	161
6.4.	Conclusions	165
7.	Weak lensing estimation with convolutional neural networks	167
7.1.	Introduction	167
7.2.	Theory	169
7.2.1.	Neural networks	169
7.2.1.1.	Training process	170
7.2.2.	Convolutional neural networks	171
7.3.	Method	173
7.3.1.	Data generation	173
7.3.2.	Model building and training	175
7.4.	Models explored and results with increasing complexity	176
7.4.1.	Complexity level 1: Same CMB realisation, different lensing realisations	177
7.4.2.	Complexity level 2: Different CMB realisations	180
7.4.3.	Complexity level 3: Reduced lensing exaggeration	183
7.4.4.	Complexity level 4: No lensing exaggeration	187
7.4.5.	Complexity level 5: Higher resolution, small field of view	190
7.4.6.	Complexity level 6: Higher resolution, wide field of view	195
7.5.	Discussion	199
8.	Conclusions	201
8.1.	Summary	201

8.2. Future prospects	202
A. Proofs supplementing exact pseudo-C_ℓ likelihood	205
A.1. General pseudo- $a_{\ell m}$ covariance derivation	205
A.2. Equivalence of the two forms of the joint characteristic function	208
Bibliography	211

Word count: 68,510

Abstract

There are many open questions in cosmology, chief among which are the nature of the dominant components in the Universe: dark energy and dark matter. One promising probe with which to make progress in answering these questions is the analysis of weak gravitational lensing: subtle distortions in the shapes of distant galaxies due to the gravitational effect of large-scale structure in the Universe. These distortions depend closely on the properties of dark energy and dark matter, which govern the evolution of structure in recent times. Careful statistical analysis of the distortions may therefore place tight constraints on physical theories of these components, along with gravity and other constituents of the Universe such as neutrinos. This promise is set to be realised by the upcoming generation of weak lensing experiments such as the *Euclid* space mission and the Square Kilometre Array radio observatory, which will observe tens of billions of galaxies, and in doing so will achieve an unprecedented level of statistical precision on cosmological constraints. However, such unprecedented precision requires equally unprecedented understanding and control of all aspects of the analysis process, in order to obtain reliable results and avoid undiagnosed biases and systematic errors.

This thesis makes progress towards a complete and robust understanding of certain aspects of weak lensing analyses. Chapters 3–6 focus on pseudo- C_ℓ estimators, which are fast estimators of two-point correlation in Fourier space, for use with partial-sky observations. The exact joint likelihood of pseudo- C_ℓ estimates from an arbitrary number of correlated spin-0 and spin-2 Gaussian fields is derived and validated in Chapter 3. It is shown in Chapter 4 that to obtain accurate constraints on dark energy parameters with pseudo- C_ℓ estimates from *Euclid*, a Gaussian likelihood is sufficiently accurate, and that this accuracy is robust to the details of the analysis setup. A Gaussian likelihood requires a covariance matrix, and in Chapter 5 a method is presented with which to calculate a complete covariance matrix of pseudo- C_ℓ estimates for *Euclid*, including non-Gaussian mode coupling arising from non-linear structure growth as well as Gaussian mode coupling arising from the convolution of the signal with the mask describing the details of the sky coverage. The resulting covariance matrix is compared to one estimated from weak lensing simulations, with good agreement. Chapter 6 turns to the question of how to select an angular binning strategy to strike an optimal balance between statistical constraining power and data compression. Finally, Chapter 7 considers a different question, of whether convolutional neural networks may be used to estimate weak lensing shear directly from radio visibilities from the Square Kilometre Array. Working towards this aim from a simplified case of lensing of the cosmic microwave background, it shows that this method is promising but also entails many challenges. The work presented in this thesis helps to make significant progress towards an ultimate goal of reliable cosmological inference from future weak lensing data, but many challenges and open questions remain, which are discussed in Chapter 8.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and has given the University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy¹, in any relevant Thesis restriction declarations deposited in the University Library, the University Library's regulations² and in the University's policy on Presentation of Theses.

¹ <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420>

² <http://www.library.manchester.ac.uk/about/regulations>

Acknowledgements

My thanks, above all, go to Michael and Lee for your supervision and guidance, and to my partner Michaela for your unwavering support. I additionally wish to thank everyone who I had the pleasure of working with within Jodrell Bank Centre for Astrophysics and the Euclid Consortium.

Discussions with many other researchers contributed significantly to the work presented in this thesis. In particular, thanks to Andrew Jaffe, Benjamin Joachimi and Andy Taylor for Chapter 3, and to Alex Hall and Peter Taylor for Chapter 4. Chapter 5 is adapted from a *Euclid* collaboration paper with too many co-authors to personally thank, many of whom contributed valuable feedback to the draft. I am especially grateful to Alan Heavens and Keith Grainge for examining the thesis as a whole and providing much valuable feedback. Chapters 3–5 also benefited from anonymous peer reviews.

The work in Chapters 4 and 5 would not have been possible without the public simulation suites of the SLICS (Harnois-Déraps & Van Waerbeke 2015; Harnois-Déraps et al. 2018) and Takahashi et al. (2017). A host of open-source code published by members of the astronomy and cosmology community was invaluable throughout this work: Astropy (The Astropy Collaboration et al. 2013, 2018), CAMB (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012), the Core Cosmology Library (Chisari et al. 2019), CosmoLike/CosmoCov (Krause & Eifler 2017; Fang, Eifler & Krause 2020), CosmoSIS (Zuntz et al. 2015), FITSIO (Pence 1999), HEALPix/healpy (Górski et al. 2005; Zonca et al. 2019), LensTools (Petri, Haiman & May 2016), NaMaster (Alonso, Sanchez & Slosar 2019), TreeCorr (Jarvis, Bernstein & Jain 2004), and Wigner (Tessore 2019). The work in this thesis has also made heavy use of a multitude of open-source Python packages: most importantly NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), and Matplotlib (Hunter 2007), supplemented by the Non-Parametric Entropy Estimation Toolbox (Ver Steeg 2014), pandas (McKinney 2010; The pandas Development Team 2020), parse (Jones et al. 2021), py3nj (Fuji 2021), quaternion (Boyle et al. 2022), Scikit-learn (Pedregosa et al. 2011), seaborn (Waskom 2021), SHTools (Wieczorek & Meschede 2018), Spherical Functions (Boyle, Stein & Gross 2021), SymPy (Meurer et al. 2017), and TensorFlow/Keras (Abadi et al. 2016; Chollet et al. 2015).

Finally, none of the work in this thesis would have been possible without the financial support of a studentship from the UK Science & Technology Facilities Council.

Supporting publications

Exact joint likelihood of pseudo- C_ℓ estimates from correlated Gaussian cosmological fields

Upham, R. E., Whittaker, L. & Brown, M. L. 2020

Monthly Notices of the Royal Astronomical Society 491, 3165

- Chapter 3 is adapted from this publication.

Sufficiency of a Gaussian power spectrum likelihood for accurate cosmology from upcoming weak lensing surveys

Upham, R. E., Brown, M. L. & Whittaker, L. 2021

Monthly Notices of the Royal Astronomical Society 503, 1999

- Chapter 4 is adapted from this publication.

Euclid: Covariance of weak lensing pseudo- C_ℓ estimates. Calculation, comparison to simulations, and dependence on survey geometry

Upham, R. E., Brown, M. L., Whittaker, L., et al. 2022

Astronomy & Astrophysics 660, A114

- Chapter 5 is adapted from this publication.

Supporting code and data

Code is made available at <https://github.com/robinupham>. The repository corresponding to each chapter is listed below.

- Chapter 3: https://github.com/robinupham/pseudo_cl_likelihood;
- Chapter 4: https://github.com/robinupham/gaussian_cl_likelihood;
- Chapter 5:
 - https://github.com/robinupham/CosmoCov_ClCov
for non-Gaussian covariance;
 - https://github.com/robinupham/shear_pcl_cov
for Gaussian covariance, plotting, and other utilities;
- Chapter 6: https://github.com/robinupham/angular_binning;
- Chapter 7: https://github.com/robinupham/cnn_lensing.

In addition, the following data repositories are available at Zenodo.

- Power spectra estimated from Gaussian field simulations used in Chapter 4:
<https://doi.org/10.5281/zenodo.4316732>;
- Power spectra estimated from Takahashi et al. (2017) weak lensing simulations, and connected non-Gaussian covariance matrix, used in Chapter 5:
<https://doi.org/10.5281/zenodo.5163132>.

Chapter 1

Cosmology

Cosmology is the study of the Universe, the components that make it up, and its past and future evolution.

This chapter begins with an overview of the current standard model of cosmology in Section 1.1, before outlining some of the key questions that remain unanswered in Section 1.2. A description of the main observational probes for learning about cosmology is given in Section 1.3, before Section 1.4 introduces the main such probe considered in this thesis: weak gravitational lensing.

1.1. Standard cosmological model

The concept of a standard model of cosmology refers to the simplest model that is consistent with all observational data to date.

The most minimal version of the standard cosmological model is called Λ CDM, which stands for its two largest constituents: dark energy in the form of a cosmological constant (Λ) and cold dark matter (CDM). These will be described in more detail in Section 1.1.4. This model has six free parameters, which are described in Section 1.1.7, but is also extendable with additional parameters. For example, time-varying dark energy may be included using the dark energy equation of state parameter w (see Section 1.1.4.3), in which case the model is sometimes known as w CDM.

1.1.1. Cosmological principle and anisotropies

The most fundamental aspect of the standard cosmological model is the cosmological principle, which states that on sufficiently large scales, the Universe is homogeneous and isotropic, meaning that it is the same everywhere and in all directions.

However, the Universe is not completely isotropic. This is self-evident on smaller scales, where we observe structure such as planets, stars and galaxies. It is also the case on larger scales, where matter is distributed in the walls, filaments and voids that make up the ‘cosmic web’, as seen in Figure 1.1.

These anisotropies are believed to arise from quantum fluctuations in the early Universe, which later resulted in over- and underdense regions of matter, which evolved under grav-

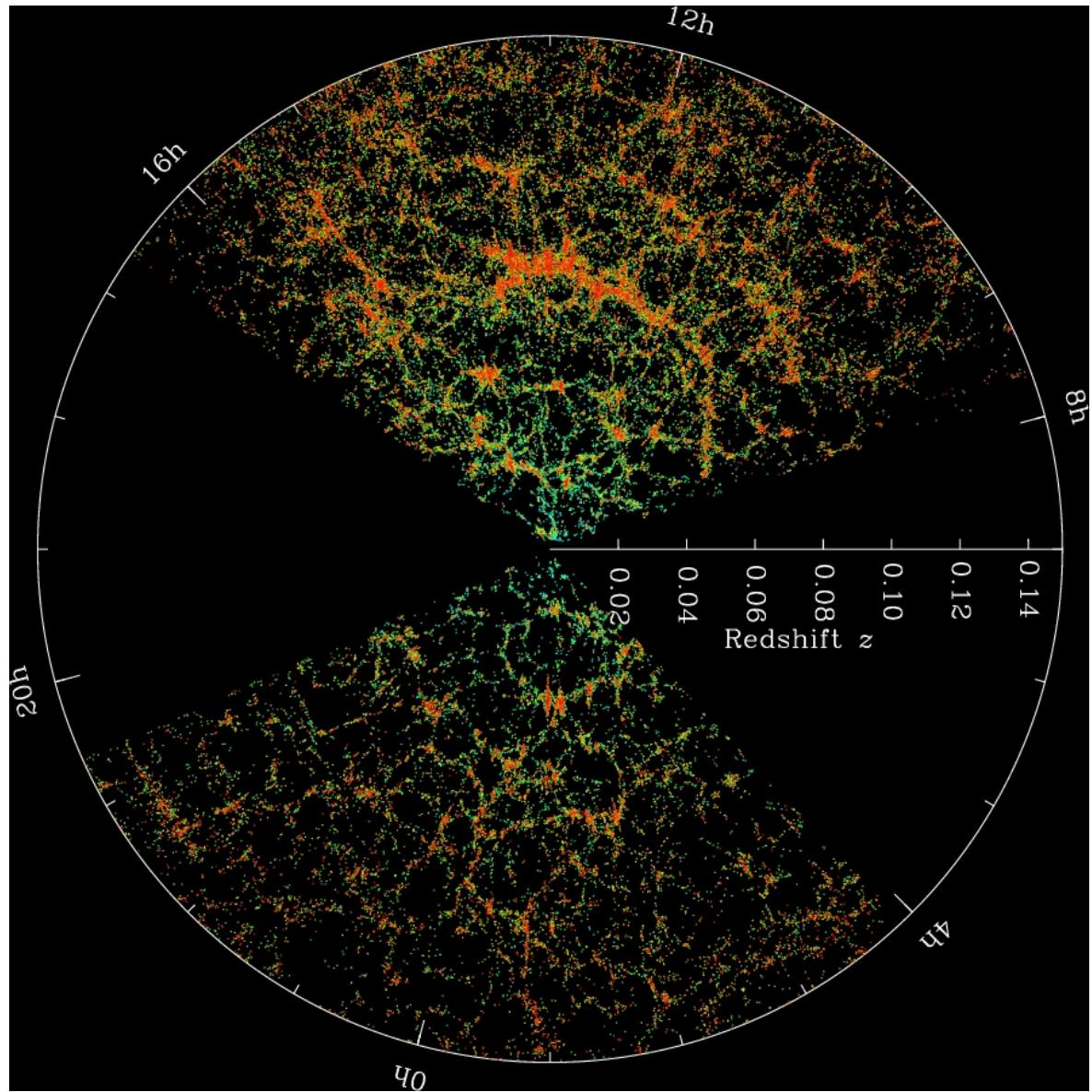


Figure 1.1. The cosmic web of the large-scale matter distribution, as traced by galaxies detected by the Sloan Digital Sky Survey (SDSS; York et al. 2000; Blanton et al. 2017). Image by M. Blanton and SDSS.

ity to give the structures we observe today. This is strongly supported by measurements of baryon acoustic oscillations (which will be described in Section 1.3.4) and by observations of the cosmic microwave background (CMB), which have revealed that the early Universe was extremely smooth, with small anisotropies of order one part in 10^5 . The CMB and its anisotropies will be described in more detail in Section 1.3.1.

1.1.2. Cosmic expansion and redshift

After the cosmological principle, perhaps the most fundamental aspect of the standard model of the Universe is that it is expanding. Mathematically, this can be described simply using the scale factor, $a(t)$, where any distance d at time t is given by

$$d(t) = a(t) d_0, \quad (1.1)$$

where d_0 is the distance today ($t = t_0$), and the current value of the scale factor, a_0 , is defined as

$$a_0 = 1. \quad (1.2)$$

For any time in the past, $t < t_0$,

$$a(t < t_0) < 1, \quad (1.3)$$

and therefore distances were shorter than they are today. Rewound far enough, this implies that at one stage the Universe was infinitely dense. This is called the Big Bang, and is a central part of the standard cosmological model.

If distances were shorter in the past, then so too were wavelengths of light and other radiation. The wavelength of any photon has subsequently been stretched between its emission in the past and its observation today. This effect is called redshift. Since the finite speed of light means that observations of any distant object are equivalent to looking back in time, redshift is a useful measure of the distance of objects from the Earth. Redshift, z , can be quantified as

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{1}{a(t)}, \quad (1.4)$$

where λ_{emit} and λ_{obs} are the emitted and observed wavelengths of light.

Evidence for the expanding Universe comes from the Hubble law (Lemaître 1927; Hubble 1929), which states that distant galaxies are receding at a speed v proportional to their distance d :

$$v = H_0 d, \quad (1.5)$$

where the constant of proportionality is the Hubble constant H_0 . Despite its name, H_0 is today considered as the current value of a time-varying parameter $H(t)$, which is related to the scale factor a as

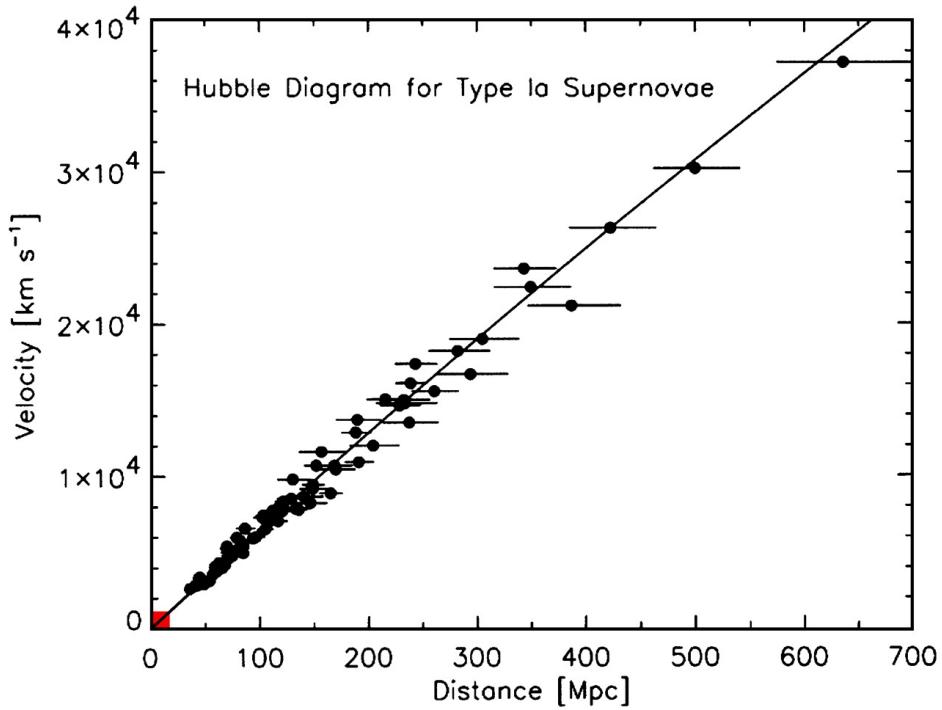


Figure 1.2. Diagram of the Hubble law (Equation 1.5) relating the recession velocity of distant objects (in this case Type Ia supernovae, which are described in Section 1.3.3) to their distance. Taken from Kirshner (2004).

$$H = \frac{\dot{a}}{a}, \quad (1.6)$$

where the overdot denotes a time derivative. A more recent diagram of the Hubble law is shown in Figure 1.2. This was compiled from Type Ia supernovae, which will be described in Section 1.3.3, and is taken from Kirshner (2004). Strong evidence for the implied Big Bang is supplied by observations of the CMB (Section 1.3.1).

1.1.3. The FLRW Universe

This section provides an overview of some of the mathematics underpinning the standard model of cosmology. It is named after four of its key contributors: Friedmann, Lemaître, Robertson and Walker.

In Λ CDM and simple extensions such as w CDM, the Universe is governed by the theory of general relativity (GR), which describes the relationship between the geometry of the Universe and its contents. GR provides a near-universal theory of gravity covering both everyday and cosmological scales (though it notably fails on quantum scales). GR is summarised in the Einstein field equations, which may in fact be written as a single equation (Einstein 1916),

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \quad (1.7)$$

$G_{\mu\nu}$ in Equation (1.7) is the Einstein tensor describing the curvature of spacetime, and $T_{\mu\nu}$ is the energy-momentum tensor describing the energy density at a given point in spacetime. $g_{\mu\nu}$ is the metric tensor, which describes the geometric and causal structure of spacetime, and is related to the separation between points in spacetime, which is discussed below. In Equation (1.7) it multiplies the cosmological constant Λ , which could alternatively be absorbed into $T_{\mu\nu}$, where it could also be replaced with a time-varying dark energy contribution to the density. The Einstein field equations are related to the Newtonian theory of gravity through the classical gravitational constant G .

Distances between points in a homogeneous, isotropic and expanding Universe can be quantified using the FLRW metric, which decomposes the line element of spacetime ds into contributions from time dt and space, $d\Sigma$,

$$ds^2 = -c^2 dt^2 + a(t)^2 d\Sigma^2. \quad (1.8)$$

The spatial metric is further decomposed, in hyperspherical coordinates, into a radial contribution $d\chi$ and an angular contribution $d\Omega$ as

$$d\Sigma^2 = d\chi^2 + S_K(\chi)^2 d\Omega^2. \quad (1.9)$$

χ is the comoving angular distance, defined in terms of the scale factor a as

$$\chi(a) = c \int_a^1 \frac{da'}{a'^2 H(a')}. \quad (1.10)$$

The angular line element $d\Omega$ is given in terms of polar and azimuthal contributions $d\theta$ and $d\phi$ as

$$d\Omega^2 = d\theta^2 + \sin^2(\theta) d\phi^2. \quad (1.11)$$

$S_K(\chi)$ in Equation (1.9) is the comoving angular diameter distance, whose value depends on the curvature of the Universe K . Specifically, it takes different forms in three cases: a closed Universe with spherical geometry ($K > 0$), a flat Universe with Euclidean geometry ($K = 0$), or an open Universe with hyperbolic geometry ($K < 0$). The comoving angular diameter distance in each of these cases is given by

$$S_K(\chi) = \begin{cases} K^{-\frac{1}{2}} \sin(K^{\frac{1}{2}}\chi) & \text{for } K > 0 \text{ (closed);} \\ \chi & \text{for } K = 0 \text{ (flat);} \\ |K|^{-\frac{1}{2}} \sinh(|K|^{\frac{1}{2}}\chi) & \text{for } K < 0 \text{ (open).} \end{cases} \quad (1.12)$$

The spacetime line element in the FLRW metric, ds , can be related to the metric tensor $g_{\mu\nu}$ from Equation (1.7) as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (1.13)$$

where dx^μ is the infinitesimal displacement in comoving coordinates; that is, spacetime coordinates defined such that the spatial component remains constant in the expanding Universe.

It is typical to model the constituents of the Universe as perfect fluids, which are completely described by their energy density ρ and isotropic pressure P . Under this assumption, a solution to the Einstein field equations for a homogeneous, isotropic and expanding Universe governed by the FLRW metric is given by the Friedmann equations (Friedmann 1922, 1924),

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda c^2}{3} - \frac{Kc^2}{a^2}; \quad (1.14)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right) + \frac{\Lambda c^2}{3}. \quad (1.15)$$

These two equations can be used to derive a third, the energy conservation equation:

$$\dot{\rho} = -3H\left(\rho + \frac{P}{c^2}\right). \quad (1.16)$$

Under the assumption that the constituents of the Universe may be modelled as perfect fluids, the equation of state relating the density ρ_X and pressure P_X for a given constituent X is

$$P_X = w_X\rho_Xc^2, \quad (1.17)$$

where w_X is the dimensionless equation of state parameter. In this thesis, the symbol w will be used without any subscript to denote the equation of state parameter for dark energy. This parameter is discussed further in Section 1.1.4.3.

The energy conservation equation (Equation 1.16) may be used to derive an expression for the evolution of the density of constituent X:

$$\frac{d \ln \rho_X}{d \ln a} + 3(1 + w_X) = 0. \quad (1.18)$$

When w_X is constant, this equation has a solution,

$$\rho_X \propto a^{-3(1+w_X)}. \quad (1.19)$$

When a given component is dominant, the resulting evolution of the scale factor as a function of time may be found by substituting ρ_X into Equation (1.14). Values of w_X for the different constituents of the Universe and the resulting evolution of the scale factor will be discussed in Section 1.1.5.

1.1.4. Constituents of the Universe

1.1.4.1. Baryonic matter

Baryonic matter is the ordinary matter that makes up what we see on Earth and within the solar system. The term is commonly used to include not only baryons—particles composed of three quarks, including protons, neutrons, and their higher-energy counterparts—but more generally any form of matter as described by the standard model of particle physics. Today it constitutes around 4.9% of the total energy density of the Universe (Planck Collaboration et al. 2020b). Compared to dark matter and dark energy, baryonic matter is much better understood. Countless high-precision tests of the standard model have been carried out at particle colliders and elsewhere, with no significant deviations from the theoretical predictions yet detected (e.g. Erler & Schott 2019, for a review).

However, there are many contradictions between the predictions of the standard model and observations of the Universe. For example, in addition to its failure to describe dark matter and dark energy (and gravity), it predicts that there should be an equal amount of matter and antimatter, which is not observed. Discrepancies such as these continue to motivate a large amount of active research in particle physics.

1.1.4.2. Dark matter

Several independent forms of observational evidence point towards the existence of some additional form of matter, constituting around 25.9% of the energy density of the Universe (Planck Collaboration et al. 2020b). This has come to be known as dark matter, since it does not interact electromagnetically and cannot be ‘seen’ directly in the traditional sense. It has only been observed to interact gravitationally, although many theories predict some small amount of interaction by other means, such as via the weak interaction.

The first evidence for dark matter arrived in the form of an excess in the observed orbital velocities of stars in nearby galaxies (Kapteyn 1922), and then galaxies themselves in clusters (Zwicky 1933, 1937), compared to the velocities that could be explained by the amount of visible matter. Without an alternative theory of gravity, these observed galaxy rotation curves could only be explained by an invisible form of matter—roughly five times as much as the visible, baryonic matter. An example of a galaxy rotation curve is shown in Figure 1.3. The data points can only be explained by combining the visible ‘disk’ component with an invisible dark matter ‘halo’.

More direct evidence for dark matter has been provided by weak lensing mass reconstructions of galaxy cluster collisions, such as those of the Bullet Cluster. The reconstructed matter distribution of this cluster was revealed to extend far beyond its visible limits, and furthermore is inconsistent with a simple modification of gravity at a significance of 8σ (Clowe, Gonzalez & Markevitch 2004; Clowe et al. 2006). The observations of the Bullet

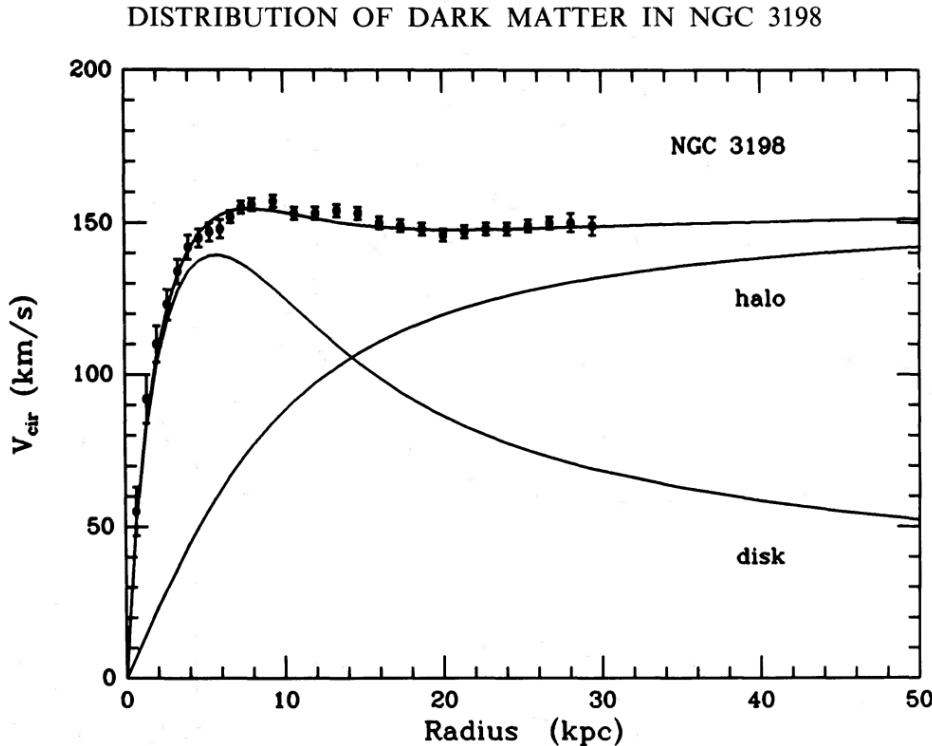


Figure 1.3. A galaxy rotation curve taken from van Albada et al. (1985). The data points can only be explained by combining the visible ‘disk’ component with an invisible dark matter ‘halo’.

Cluster also demonstrate the collisionless nature of dark matter (Markevitch et al. 2004).

Further evidence for the existence of dark matter comes from observations of the CMB. Around six times the amount of matter that we observe directly is needed to achieve a good fit to the Λ CDM model (Planck Collaboration et al. 2020b). The CMB will be discussed further in Section 1.3.1.

Dark matter is required to be ‘cold’, i.e. sufficiently massive as to move non-relativistically, in order to explain the observed level of structure in the present-day Universe (Bond, Szalay & Turner 1982; Blumenthal, Pagels & Primack 1982; Peebles 1982; Blumenthal et al. 1984; Davis et al. 1985).

However, despite the relative abundance of observational evidence for dark matter, a direct detection on Earth continues to be elusive. Collider experiments have performed extensive searches over the past decade, and classes of physical models of dark matter that were not long ago believed to be likely have now been largely ruled out (Trevisani 2018; Aaboud et al. 2018; ATLAS Collaboration 2019).

1.1.4.3. Dark energy

Almost all of the remaining energy density in the Universe—around 69.1% (Planck Collaboration et al. 2020b)—is believed to be some additional unknown form of energy responsible for the accelerating expansion of the Universe. This additional form of energy is known as dark energy.

1.1.4.3.1. Evidence for dark energy

The main evidence for the existence of dark energy comes from the accelerating expansion of the Universe, which was discovered by precision observations of distant Type Ia supernovae in the late 1990s (Riess et al. 1998; Perlmutter et al. 1999). Type Ia supernovae will be discussed further in Section 1.3.3. This accelerating expansion cannot be explained under GR without a large additional and previously unknown contribution to the total energy density.

Measurements of CMB anisotropies indicate that the Universe is flat or almost flat (Planck Collaboration et al. 2020b). Under the Λ CDM model, this cannot be explained by matter alone, including dark matter, and therefore requires a large contribution from dark energy. Probes of the large scale structure of the Universe, such as weak lensing, also detect too little matter for flatness (e.g. DES Collaboration et al. 2022).

1.1.4.3.2. Models of dark energy

Cosmological constant

The mathematically simplest way to account for an additional energy source in GR and its solutions is with a cosmological constant, Λ . The cosmological constant was first introduced by Einstein in his original descriptions of GR (Einstein 1917) in order to provide a steady state Universe, which was generally assumed to be the case at the time, when his equations otherwise predicted a dynamic Universe. After the Universe was discovered to in fact be expanding, the cosmological constant was assumed to be zero, and Einstein is alleged to have described its initial inclusion as his “biggest blunder” (O’Raifeartaigh & Mitton 2018). However, after the accelerating expansion was discovered, the cosmological constant was reinstated and today forms an essential pillar of the Λ CDM standard model of cosmology.

The cosmological constant gives a dark energy equation of state parameter w of

$$w = -1. \tag{1.20}$$

Physically a cosmological constant corresponds to an intrinsic vacuum energy of space. Such a vacuum energy is predicted by the standard model of particle physics, but is around

120 orders of magnitude too large to explain cosmological observations (e.g. Adler, Casey & Jacob 1995). This is known as the cosmological constant problem, and is one of the major outstanding questions in fundamental physics.

Quintessence

The term quintessence refers to a proposed new scalar field, many plausible models of which exist (see Tsujikawa 2013 for a review). Such a field could change over time, such that the dark energy equation of state parameter w is a function of the scale factor a ,

$$w = w(a). \quad (1.21)$$

It is common to Taylor expand Equation (1.21) to linear order about $a = 1$ to define w_0 and w_a as

$$w(a) = w_0 + w_a(1 - a), \quad (1.22)$$

such that

$$w_0 = w(a = 1); \quad (1.23)$$

$$w_a = -\frac{dw}{da}(a = 1). \quad (1.24)$$

Observations to date are consistent with the values of

$$w_0 = -1; \quad (1.25)$$

$$w_a = 0 \quad (1.26)$$

(e.g. Ribeiro 2019, from a combined analysis of CMB anisotropies, baryon acoustic oscillations, supernovae and cosmic chronometers), which corresponds to a cosmological constant.

Modified gravity

It is possible to negate the need for dark energy altogether with suitable modifications to gravity, which deviate from GR (e.g. Nojiri & Odintsov 2003, 2011; Nicolis, Rattazzi & Trincherini 2009; Clifton et al. 2012; Nojiri, Odintsov & Oikonomou 2017). Alternatively, dark energy could exist alongside a modified theory of gravity. However, all observations to date are consistent with GR, provided that dark energy is included, and some modified gravity models have been ruled out by recent observations of gravitational waves (Blas et al. 2016; Vainio & Vilja 2017; Arai & Nishizawa 2018; Battye, Pace & Trinh 2018; Ma & Yunes 2019). Pulsar timing experiments also place tight constraints on deviations from GR (Beltrán Jiménez, Piazza & Velten 2016; Shao et al. 2017; Cai et al. 2019; Kramer et al. 2021).

1.1.4.4. Other constituents

The Universe also contains small amounts of ‘radiation’, totalling less than 1% of the total energy density. This includes neutrinos, and the CMB, which will be described further in Section 1.3.1.

1.1.5. Expansion history

The Universe today is dominated by dark energy, but this is a fairly recent development. Prior to this, the Universe is believed to have gone through several epochs in which different components were dominant, each of which will now be described.

1.1.5.1. Inflation

It is believed that the early Universe underwent a short period of rapid expansion, known as inflation (Guth 1981; Guth & Pi 1982; Starobinsky 1982; Linde 1982, 1983). The inflationary period is thought to have lasted from 10^{-36} s to 10^{-33} – 10^{-32} s after the Big Bang, during which the Universe expanded by around 60 *e-folds* (a factor of 10^{26}) (Planck Collaboration et al. 2020c). Inflation provides a mechanism to create the seeds for the present-day large-scale structure of the Universe, by vastly magnifying quantum fluctuations in the early Universe. The popularity of the theory is also driven by its ability to solve two significant cosmological problems: the horizon problem and the flatness problem.

The horizon problem is the fact that the CMB is isotropic to within one part in 10^5 , implying that regions far apart on the sky must have been in thermal equilibrium in the early Universe, when—in the absence of inflation—they could have never been in causal contact. Invoking inflation allows seemingly distant parts of the CMB sky to have been in causal contact prior to inflationary expansion.

The flatness problem arises from the observation that the present-day Universe is extremely close to flat (Planck Collaboration et al. 2020b). It can be shown from Equations (1.14)–(1.16) that flatness is an unstable equilibrium, in that any perturbation from flatness should grow over time. Therefore, the fact that the Universe is extremely close to flat today implies that any deviation from flatness in the early Universe must have been infinitesimal—of order 10^{-55} or smaller (Guth 1981). The lack of an explanation—in the absence of inflation—for this suspiciously convenient value is a type of ‘fine-tuning’ problem. Inflation provides an explanation, because such a rapid expansion of the Universe strongly suppresses any previous cosmic curvature and leaves a Universe sufficiently flat to explain current observations.

Inflation is also able to predict the observed value of the scalar spectral index n_s (Bardeen, Steinhardt & Turner 1983; Planck Collaboration et al. 2020c). (The scalar spectral index and

other parameters of the Λ CDM model are discussed in Section 1.1.7.) Physically, inflation can be explained by a new scalar ‘inflaton’ field, many plausible models of which exist (e.g. Baumann & McAllister 2015, for a review). Future CMB polarisation experiments hope to detect direct evidence of the inflaton field, which would confirm the theory of inflation and constrain physical models (see Section 1.3.1).

1.1.5.2. Radiation domination

The Universe subsequently underwent a period of radiation domination, during which the dominant components were relativistic photons and neutrinos. Radiation is modelled as a gas, exerting a positive pressure in each of three spatial dimensions, giving it an equation of state parameter w_r of

$$w_r = \frac{1}{3}. \quad (1.27)$$

Inserting this into Equation (1.19) gives the evolution of the radiation density ρ_r as

$$\rho_r \propto a^{-4}. \quad (1.28)$$

When radiation is dominant, the first Friedmann equation (Equation 1.14) then becomes

$$H^2 \propto a^{-4}. \quad (1.29)$$

Using the definition of the Hubble parameter H in terms of the time derivative of the scale factor (Equation 1.6), we can obtain a solution for the scale factor as a function of time in the epoch of radiation domination,

$$a \propto t^{1/2}. \quad (1.30)$$

1.1.5.3. Matter domination

The period of radiation domination was followed by a period of matter domination. Matter is pressureless, and therefore its equation of state parameter w_m is

$$w_m = 0, \quad (1.31)$$

giving the evolution of the matter density ρ_m as

$$\rho_m \propto a^{-3}. \quad (1.32)$$

Comparing this to Equation (1.28) for radiation explains why matter eventually came to dominate over radiation, since radiation was more heavily diluted by the expansion of the Universe. This effect can be understood in terms of redshift: as the Universe expanded, radiation is redshifted, decreasing its energy by an additional factor $1/a$.

In the epoch of matter domination, the Hubble parameter evolves as

$$H^2 \propto a^{-3}, \quad (1.33)$$

which gives the evolution of the scale factor as

$$a \propto t^{2/3}. \quad (1.34)$$

Despite the faster growth of the scale factor during matter domination, it was in this period that structure was most able to form. Structure formation was strongly suppressed in the period of radiation domination by the high density of relativistic particles.

1.1.5.4. Dark energy domination

The Universe is now in a period of dark energy domination. The dark energy equation of state parameter w is either exactly or approximately equal to -1 , depending on the model (see Section 1.1.4.3). Inserting a value of $w = -1$, corresponding to a cosmological constant Λ , into Equation (1.19) gives

$$\rho_\Lambda \propto 1, \quad (1.35)$$

i.e. a constant density. This is how dark energy has come to eventually dominate, as the density of matter and radiation is diluted by the expanding Universe. For dark energy domination, Equation (1.14) gives that

$$H^2 = \frac{\Lambda c^2}{3}, \quad (1.36)$$

which has an exponential solution for the scale factor:

$$a \propto \exp \left[\sqrt{\frac{\Lambda}{3}} ct \right]. \quad (1.37)$$

Under dark energy domination, therefore, the Universe undergoes exponential growth, which suppresses structure formation. This means that the growth of structure as a function of redshift z (or equivalently, the scale factor a) depends heavily on the nature of dark energy and its potential time evolution. Observational probes of structure growth and the recent expansion history of the Universe are therefore able to potentially place tight constraints on dark energy models via their predictions of $w(a)$.

1.1.6. Density parameters

It is convenient to define the critical density ρ_c , as the density required for a flat Universe, obtained by setting $K = 0$ and $\Lambda = 0$ in Equation (1.14):

$$\rho_c = \frac{3H^2}{8\pi G}. \quad (1.38)$$

For each component X of the Universe, we may then define the density parameter Ω_X ,

$$\Omega_X = \frac{\rho_X}{\rho_c}. \quad (1.39)$$

We may include the curvature contribution to the Friedmann equation in this form too by defining it as a density ρ_K ,

$$\rho_K = -\frac{3Kc^2}{8\pi Ga^2}, \quad (1.40)$$

giving the curvature density parameter Ω_K ,

$$\Omega_K = \frac{\rho_K}{\rho_c} = -\frac{Kc^2}{a^2 H^2}. \quad (1.41)$$

Including a general dark energy contribution, the Friedmann equation (Equation 1.14) may then be written in form which makes explicit the way in which the different components evolve over time:

$$\frac{H^2}{H_0^2} = \frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} + \frac{\Omega_K}{a^2} + \frac{\Omega_{DE}}{a^{3(1+w)}}. \quad (1.42)$$

If dark energy is a cosmological constant, then the last term in Equation (1.42) becomes simply Ω_Λ , with

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H^2}. \quad (1.43)$$

1.1.7. Free parameters

The minimal Λ CDM model has six free parameters. There is not a single unique parameterisation, but a typical choice is the following (e.g. Di Valentino, Melchiorri & Silk 2015):

- The Hubble parameter H_0 , which may alternatively be described using the dimensionless Hubble parameter h ,

$$h = \frac{H_0}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}}. \quad (1.44)$$

- $\Omega_b h^2$, where Ω_b is the energy density of baryons as a fraction of the critical density.
- $\Omega_c h^2$, where Ω_c is the dark matter energy density as a fraction of the critical density.
- The amplitude of primordial scalar perturbations A_s .
- The scalar spectral index n_s , describing the scale dependence of primordial perturbations.
- The optical depth to reionisation τ : a physical description of the redshift of the epoch of reionisation, which was a period in the evolution of the Universe when the first

stars and other luminous sources were formed, and were able to ionise hydrogen in the intergalactic medium.

Other parameters may be derived from these six, and vice versa. The parameters that are constrained by a particular experiment will depend on the cosmological probe in question. Two derived parameters that are of interest in this thesis are:

- Ω_m , the total matter density, given by

$$\Omega_m = \Omega_c + \Omega_b. \quad (1.45)$$

- σ_8 , the amplitude of present-day matter fluctuations on scales of $8 h^{-1} \text{Mpc}$. This may be constrained directly by probes of large scale structure, but can also be predicted by the ΛCDM model given the free parameters listed above.

The ΛCDM model may be extended by adding additional parameters. One extension is of particular interest in this thesis, which is the extension to time-varying dark energy via a dark energy equation of state parameter that depends on the scale factor as $w(a)$. As introduced in Equations (1.22)–(1.24), this may be Taylor expanded to linear order to define w_0 and w_a , which are the two cosmological parameters of greatest interest to this thesis. The extension to ΛCDM to include w_0 and w_a is often called $w\text{CDM}$.

Current constraints on the ten parameters listed above are given in Table 1.1. These are taken from Planck Collaboration et al. (2020b), where they are derived from CMB temperature, polarisation and lensing, combined with baryon acoustic oscillations and, for the dark energy parameters w_0 and w_a , Type Ia supernovae. Each of these probes is described in Section 1.3.

1.2. Outstanding questions

The previous section has described in some detail the current best understanding of the nature of the Universe. However, many questions remain outstanding. Below is a brief summary of a selection of key open questions in cosmology.

- **Nature of the constituents.** As described in Section 1.1.4, around 95% of the energy density of the Universe consists of dark energy and dark matter. However, the physical nature of both of these components remains unknown, along with an explanation for why the cosmological constant does not correspond to the vacuum energy predicted by the standard model of particle physics.
- **Modified gravity.** Is GR correct, or is a modified theory of gravity required?
- **Inflation.** What is the physical nature of the mechanism responsible for inflation?

Table 1.1. Current constraints on the ten cosmological parameters described in Section 1.1.7. Taken from Planck Collaboration et al. (2020b) Tables 2 and 6, where they are derived from CMB temperature, polarisation and lensing, combined with baryon acoustic oscillations and, for w_0 and w_a , Type Ia supernovae.

Parameter	Best-fit value \pm 68% limits
$H_0 \left[\text{km s}^{-1} \text{Mpc}^{-1} \right]$	67.66 ± 0.42
$\Omega_b h^2$	0.02242 ± 0.00014
$\Omega_c h^2$	0.11933 ± 0.00091
$10^9 A_s$	2.105 ± 0.030
n_s	0.9665 ± 0.0038
τ	0.0561 ± 0.0071
Ω_m	0.3111 ± 0.0056
σ_8	0.8102 ± 0.0060
w_0	-0.957 ± 0.080
w_a	$-0.29^{+0.32}_{-0.26}$

- **Tensions in measured parameters.** It is common for different experiments to infer values of cosmological parameters that are not in perfect agreement. However, two particular discrepancies have arisen between experiments measuring the late-time Universe and the early Universe. The most significant of these is the tension in the Hubble parameter H_0 , between local measurements of Type Ia supernovae calibrated by Cepheid variable stars using the distance ladder method (which will be described in Section 1.3.3), and values inferred from CMB observations conditioned on the Λ CDM model. This tension between early and late time measurements is illustrated in Figure 1.4. This figure is taken from Di Valentino et al. (2021), since which the tension has been claimed to have reached the 5σ significance level (Riess et al. 2022). The second tension of interest is in the parameter S_8 , defined as¹

$$S_8 = \sigma_8 \left[\frac{\Omega_m}{0.3} \right]^{0.5}, \quad (1.46)$$

between local measurements using weak lensing compared to those inferred from CMB observations. This tension is currently claimed to stand at the $2\text{--}3\sigma$ level based on one recent weak lensing analysis from the Kilo-Degree Survey (Heymans et al. 2021), although better agreement with the CMB is found by the Dark Energy Survey (DES Collaboration et al. 2022). (Both of these current weak lensing surveys

¹ The power of 0.5 in Equation (1.46) is sometimes varied slightly depending on the best fit to the data, such as to 0.48 in Rhodes, Refregier & Groth (2001).

will be described in Section 1.4.3.) While the S_8 tension may just be a statistical fluke, or the result of undiagnosed systematic errors, the H_0 tension has persisted and strengthened despite being repeatedly examined with increasing care (e.g. Riess et al. 2018a,b,c, 2020, 2021, 2022; Burns et al. 2018; Yuan et al. 2019; Reid, Pesce & Riess 2019; Macaulay et al. 2019; Soltis, Casertano & Riess 2021; Anand et al. 2022; Romaniello et al. 2022), and a resolution may require new physics (Di Valentino et al. 2021).

1.3. Observational probes

There are a number of observational methods by which our current best understanding of the Universe, as described in Section 1.1, has been developed, and using which we hope to answer some of the remaining outstanding questions such as those summarised in Section 1.2. A selection of these methods are described in this section.

1.3.1. Cosmic microwave background

Observations of the cosmic microwave background (CMB) have perhaps offered more insight into the nature of the Universe than any other method. The CMB consists of photons from the early Universe that have been redshifted over time and now reside in the microwave regime, corresponding to a temperature of around 2.7 K.

The origin of the CMB can be traced back to the first $\sim 380\,000$ years after the Big Bang, during which the Universe was sufficiently hot that it was impossible for atoms to form without being ionised, and high-energy photons existed in thermal equilibrium with protons and electrons. As the Universe expanded, it cooled, and eventually the photons had insufficient energy to ionise atoms, and the first hydrogen atoms were born. This period, called recombination, occurred at a redshift of $z \sim 1100$.

The CMB contains small anisotropies in both its temperature and polarisation, of order one part in 10^5 and 10^6 respectively, which are understood to be the precursor to the large-scale structure in the present-day Universe. (Baryon acoustic oscillations, discussed in Section 1.3.4 below, provide strong evidence for this link between the early and late time Universe.) The scale dependence of these anisotropies, captured by the power spectrum (see Chapter 2), are predicted by the Λ CDM model and depend closely on its free parameters. As a result, observations of the CMB are able to constrain many cosmological parameters, such as the amount of baryonic matter, dark matter, dark energy and curvature in the Universe, as well as its age and properties of primordial perturbations.

The CMB was proposed and discovered—initially by accident—in the mid-20th century

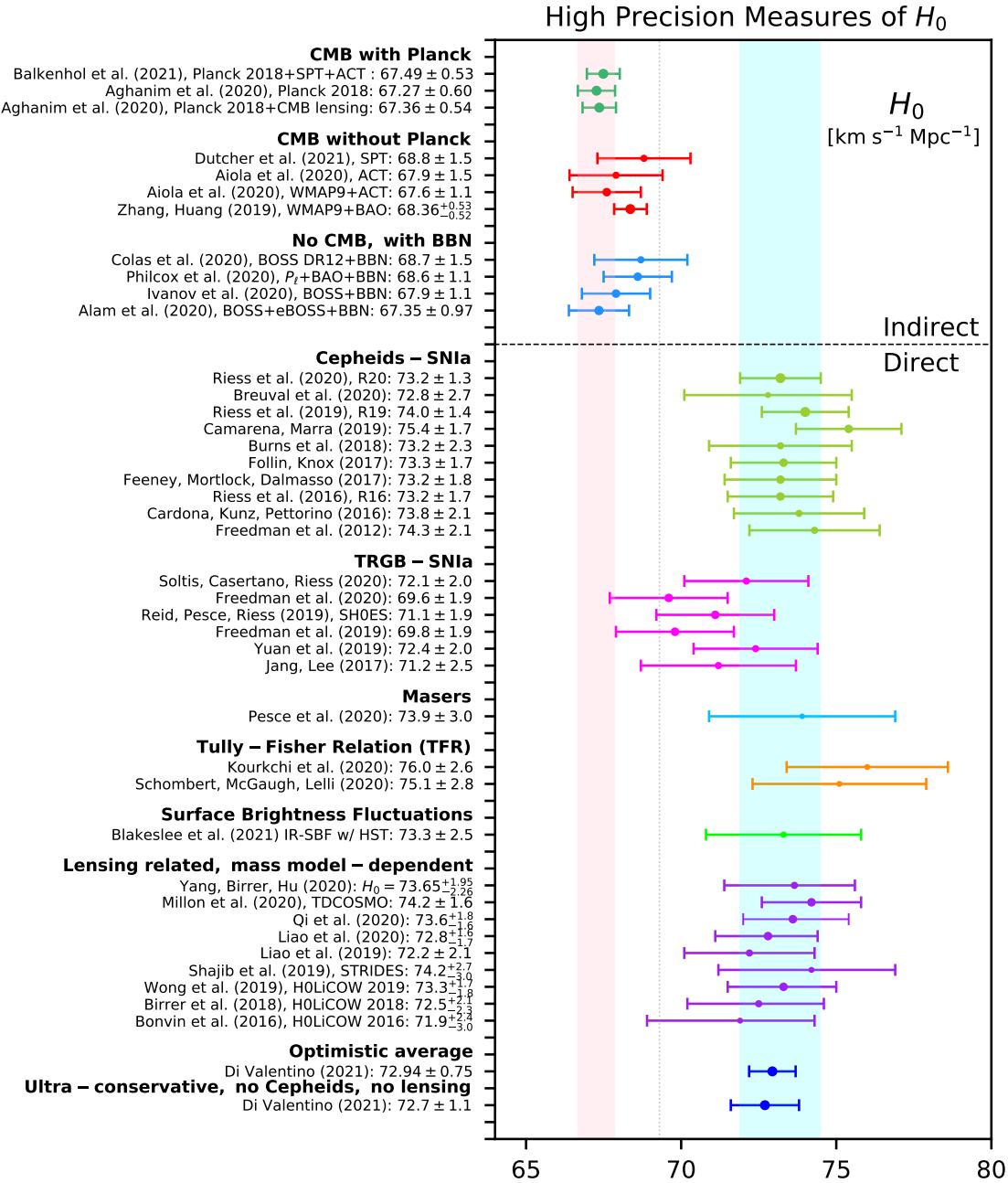


Figure 1.4. Tension in measurements of the Hubble parameter H_0 between early (indirect) and late time (direct) observations. Figure taken from Di Valentino et al. (2021). Error bars are 68% credible intervals. The blue band corresponds to the result from Riess et al. (2020) and the pink to those from Planck Collaboration et al. (2020b).

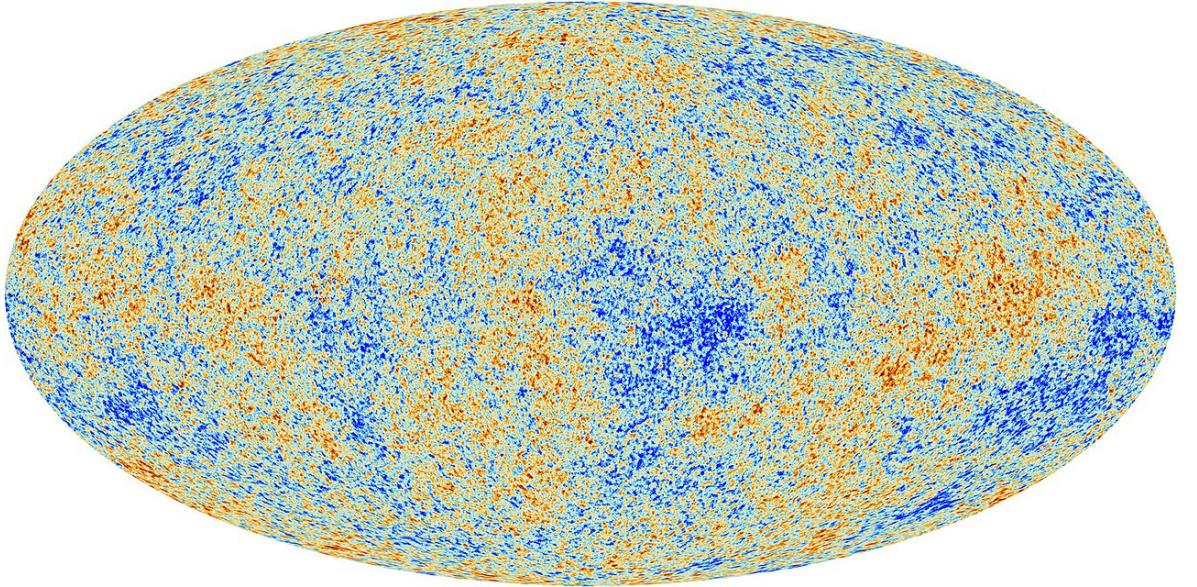


Figure 1.5. Map of CMB temperature anisotropies measured by the *Planck* satellite (Planck Collaboration et al. 2020a). Red colours are warmer than average and blue colours cooler, both by up to $300 \mu\text{K}$.

(Gamow 1948a,b; Alpher & Herman 1948a,b; Doroshkevich & Novikov 1964; Penzias & Wilson 1965; Dicke et al. 1965). A series of satellite missions in the late 20th and early 21st centuries made increasingly precise observations, free from interference from terrestrial radio sources. The first was the NASA Cosmic Background Explorer (COBE) mission, which operated from 1989 to 1993 and confirmed that the CMB radiation has an almost perfect blackbody spectrum, consistent with theoretical predictions (Fixsen et al. 1996). It was also able to measure the intrinsic anisotropies for the first time (Bennett et al. 1996). The anisotropies were measured to much higher precision with the subsequent NASA Wilkinson Microwave Anisotropy Probe (WMAP) mission, which operated from 2001 to 2010 and was able to resolve smaller scale features in the CMB temperature power spectrum. Finally, the ESA *Planck* mission, which operated from 2009 to 2013, achieved higher resolution still, as well as measuring anisotropies in the polarisation of the CMB. The map of CMB temperature anisotropies from the *Planck* mission is shown in Figure 1.5. The final cosmological analysis of *Planck* data in Planck Collaboration et al. (2020b) stands as the current pinnacle of precision cosmology, achieving sub-percent precision in many cosmological parameters.

Current and future CMB experiments such as the South Pole Telescope (Carlstrom et al. 2011), the Atacama Cosmology Telescope (Swetz et al. 2011), the POLARBEAR experiment (Kermish et al. 2012), the Simons Observatory (Ade et al. 2019), CMB-S4 (Abazajian et al. 2016) and LiteBIRD (Hazumi et al. 2019) have science goals including the detection of direct evidence of inflation via the measurement of primordial *B*-mode polarisation (Kamionkowski & Kovetz 2016) and the measurement of spectral distortions, which probe the thermal history of the Universe (Chluba et al. 2019, 2021).

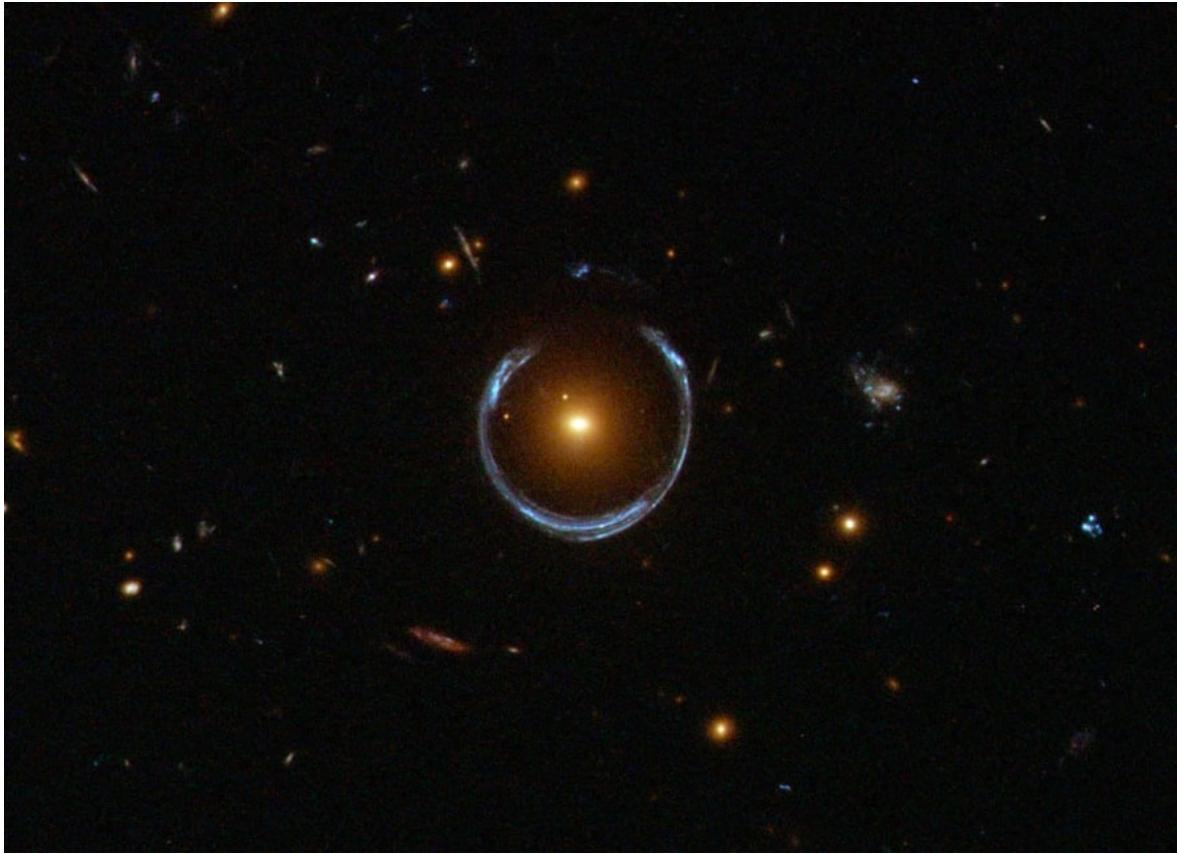


Figure 1.6. A strong lensing event captured by the Hubble Space Telescope, in which a distant blue galaxy is lensed by a foreground red galaxy. Image credit: ESA/Hubble & NASA.

1.3.2. Gravitational lensing

Gravitational lensing is the name given to the distortion of light by gravity. It was predicted by GR (Einstein 1916, 1936), in which gravity is modelled as a consequence of the distortion of spacetime by an uneven distribution of mass, which has the effect of not only bending the paths of massive objects but also of light. The prediction was confirmed soon afterwards, with measurements of the deflection of the images of stars by the gravity of the Sun during a solar eclipse (Dyson 1917; Dyson, Eddington & Davidson 1920).

Occasionally the alignment of a source and lens object is such that strong lensing occurs, in which the image of the source is distorted so strongly that the lensing effect is obvious without the need for any statistical analysis. An example is shown in Figure 1.6, in which the image of a distant blue galaxy has been strongly lensed by a foreground red galaxy. Analysis of strong lensing events like this can help to constrain models of dark matter and gravity, since the precise distortion of the source image depends sensitively on these factors (e.g. Vegetti et al. 2014; Li et al. 2016; Hezaveh et al. 2016; Gilman et al. 2020; Andrade et al. 2022). Strong lensing observations have also been used to constrain the Hubble parameter H_0 , because strong gravitational fields not only distort images but also the time taken for

light to pass through the field (e.g. Bonvin et al. 2017).

Strong lensing is rare, but essentially every line of sight on the sky is lensed weakly to some degree. This means that observations of distant galaxies are able to trace the large scale structure of the Universe through weak lensing. Weak gravitational lensing is the main observational probe of cosmology considered in this thesis, and is described in more detail in Section 1.4.

1.3.3. Type Ia supernovae

Type Ia supernovae are the end-life stage of binary star systems in which one or both stars is a white dwarf. Accretion onto the white dwarf from its companion eventually takes it beyond a certain threshold mass, which triggers runaway nuclear fusion leading to a supernova explosion.

The usefulness of Type Ia supernovae to cosmology lies in the fact that this fixed critical mass—around 1.44 solar masses—leads to all Type Ia supernovae exploding at the same peak brightness. Such an object of known brightness is known as a standard candle, and allows the calculation of the distance to the object by comparison with the apparent brightness from Earth. Although Type Ia supernovae are thought to all have the same intrinsic peak brightness, this brightness is not itself predicted by theory. This necessitates the calibration of the brightness using observations of supernovae whose distance is known by some other means. A common method involves observations of Cepheid variables in the same galaxies as the supernovae. Cepheid variables are pulsating stars, whose pulsation period is related to their brightness, such that the same method can be used to determine their distance by measuring their pulsation period. The relationship between period and brightness has to itself be calibrated, using parallax measurements of nearby Cepheids. This method of repeated calibration to obtain reliable distance measurements is known as the cosmic distance ladder.

Observations of distant Type Ia supernovae were used to detect cosmic acceleration (Riess et al. 1998; Perlmutter et al. 1999), and are still regularly used for determining the Hubble constant H_0 (Dhawan, Jha & Leibundgut 2018; Burns et al. 2018; Macaulay et al. 2019; Taubenberger et al. 2019; Khetan et al. 2021; Riess et al. 2021, 2022).

1.3.4. Baryon acoustic oscillations

Baryon acoustic oscillations are a particular scale dependence in the large-scale matter distribution in the present-day Universe. They arise from acoustic waves in the early Universe prior to recombination, which were formed from the counteracting forces of gravity and radiation pressure surrounding overdense regions of the primordial plasma. In this

period, baryons and photons were coupled, but following recombination the photons free-streamed away to form the CMB, relieving the radiation pressure and leaving the baryons essentially frozen in shells of a particular size. This size is given by the sound horizon at the time of recombination, which can be predicted by theory and has been confirmed to high precision by CMB observations (Planck Collaboration et al. 2020a). These shells evolved gravitationally to form a distinct signal in the late-time matter distribution in the Universe. Measurements of the baryon acoustic oscillation signal in large-scale structure provide a clear link between scales in the early and late Universe, while measurements of the signal as a function of redshift directly probe the expansion history of the Universe. This defined scale is known as a standard ruler, by analogy with standard candles such as Type Ia supernovae described in Section 1.3.3 above.

Baryon acoustic oscillations have been detected in the past two decades by galaxy surveys including the Sloan Digital Sky Survey (SDSS; Eisenstein et al. 2005; Percival et al. 2010), the 2dF Galaxy Redshift Survey (2dFGRS; Cole et al. 2005), the 6dF Galaxy Survey (6dFGS; Beutler et al. 2011), the WiggleZ Dark Energy Survey (Blake et al. 2011; Kazin et al. 2014) and the SDSS (extended) Baryon Oscillation Spectroscopic Survey (BOSS/eBOSS; Anderson et al. 2012, 2014; Delubac et al. 2015; de Sainte Agathe et al. 2019; Blomqvist et al. 2019).

1.3.5. Redshift-space distortions

Redshift-space distortions are distortions of galaxy positions in redshift space relative to their positions in real space, due to additional contributions to their redshift beyond the dominant Hubble expansion. These can be due to orbital Doppler shifts, or general relativistic gravitational redshift. Observations of redshift-space distortions can be used to constrain models of cosmological structure formation (Percival & White 2009; Macaulay, Wehus & Eriksen 2013; Howlett et al. 2015) as well as testing GR and constraining theories of modified gravity (Beutler et al. 2014; Percival et al. 2011; Raccanelli et al. 2013).

1.3.6. Gravitational waves

Gravitational waves are travelling distortions of spacetime emanating from accelerating masses. They are predicted by GR (Einstein 1916; Einstein 1918) but were only recently directly detected for the first time, by the Laser Interferometer Gravitational-Wave Observatory (LIGO) and the Virgo Interferometer (Abramovici et al. 1992; Accadia et al. 2012; Abbott et al. 2016). Gravitational wave detections are useful as a test of GR, and have already imposed significant constraints on modified gravity models (Blas et al. 2016; Vainio & Vilja 2017; Arai & Nishizawa 2018; Battye, Pace & Trinh 2018; Ma & Yunes 2019). Gravitational waves resulting from black hole or neutron star mergers may be used to constrain cosmic expansion if accompanied by an electromagnetic detection, since the gravitational

wave signal can be used to deduce distance while the electromagnetic signal provides redshift, thus directly constraining the Hubble constant H_0 via Equation (1.5) (Abbott et al. 2017). Future gravitational wave observatories such as the Kamioka Gravitational Wave Detector (KAGRA; Akutsu et al. 2019) and the space-based Laser Interferometer Space Antenna (LISA; Amaro-Seoane et al. 2017) may be able to probe inflation by detecting a gravitational wave background from the early Universe (Bartolo et al. 2016; Caprini et al. 2019; Maggiore et al. 2020; Kawamura et al. 2021).

1.3.7. Hydrogen 21 cm line

The Hydrogen 21 cm line is an emission line caused by a spin-flip in neutral hydrogen, between two states with an energy difference of around $5.9 \mu\text{eV}$, leading to an emission wavelength of around 21.1 cm. Since the distribution of neutral hydrogen follows the large-scale structure of the Universe, the 21 cm line can be used a probe of the large-scale structure. In a low-redshift context, this can have similar uses to other probes of large-scale structure such as galaxy surveys; for example, to constrain models of dark energy and modified gravity (Hall, Bonvin & Challinor 2013; Bull et al. 2015). However, much attention is focused on high-redshift observations of the 21 cm line, since it is the only known way to probe the ‘dark ages’ between recombination and reionisation. This is an important period of structure growth in the Universe, during which the first stars and galaxies were formed. High-redshift 21 cm observations will probe this early growth of structure, which should tightly constrain the properties of dark matter as well as potentially revealing signals from the early Universe (Furlanetto, Peng Oh & Briggs 2006; Pritchard & Loeb 2012).

Measurements of the 21 cm line are challenging due to the faintness of the signal and interference from other radio sources, both terrestrial and astrophysical. However, it has been estimated that future radio observatories such as the Square Kilometre Array (SKA; see Section 1.4.5) will be sufficiently sensitive to measure the power spectrum of the 21 cm line at the epoch of reionisation, provided foreground sources are sufficiently well understood (Pober et al. 2014).

1.4. Weak gravitational lensing

Weak gravitational lensing is the main observational probe of cosmology considered in this thesis. This section gives a qualitative description of weak lensing as a cosmological probe, while some of the more mathematical aspects are introduced in Chapter 2.

As discussed in Section 1.3.2, light paths may be distorted by passing through gravitational fields. This means that in principle, the light observed in any direction on the sky has been subtly distorted by the mass distribution between the source and the observer. This

offers particular value when applied to distant galaxies: the shape of each galaxy is subtly distorted by everything along the line of sight from the galaxy to the observer on Earth. For cosmologically distant galaxies with redshifts up to $z \sim 1\text{--}2$, there is potentially a large amount of matter along the line of sight to contribute to this distortion.

In principle, if the undistorted shape of a galaxy were known, it could be compared with the observed shape to infer the projected mass distribution to the galaxy. In practice, the intrinsic shape of a specific individual galaxy is unknown, but progress may be made with a statistical treatment of a large number of galaxies. If, to first order, we assume that galaxies are oriented randomly with no preferential alignment (which is not quite true in practice; see the section on intrinsic alignments in Section 1.4.2), then we can look out for a preferred alignment of a large number of galaxies in a particular part of the sky to uncover the projected mass distribution in that region. Repeating this process over the whole sky can probe the entire projected large-scale structure of the Universe, out to the limiting redshift of the survey. Weak gravitational lensing by the large-scale structure of the Universe is known as cosmic shear.

The scientific value of a weak lensing analysis can be increased further using the technique of tomography: splitting source galaxies into bins depending on their redshift. This gains two additional sources of information: first, redshift corresponds to distance, which introduces three-dimensional information about the large-scale structure. Second, redshift corresponds to time, which introduces information about the late-time evolution of structure. This is why weak lensing is such a promising probe of dark energy: as discussed in Section 1.1.5.4, the evolution of structure in the recent Universe depends heavily on the nature of dark energy, and specifically its equation of state parameter $w(a)$. Weak lensing also probes everything to do with matter, and therefore is a valuable tool to constrain the properties of dark matter. Its unique advantage relative to other probes of large scale structure such as galaxy clustering is that it only depends on mass distributions and the lensing theory predicted by GR, and not on additional factors such as galaxy bias.

1.4.1. Combination with galaxy clustering

It is common to combine weak lensing and galaxy clustering analyses. Galaxy clustering is the statistical analysis of the positions of galaxies on the sky, often as a function of redshift, which—similar to weak lensing—traces the large-scale structure of the late-time Universe. Since weak lensing necessarily requires measuring the shapes and positions of large numbers of galaxies, galaxy clustering may also be included in the analysis for relatively little additional work.

The advantage of combining weak lensing and galaxy clustering data lies in their respective challenges and sources of systematic error. Dominant sources of systematic uncertainty in weak lensing analyses such as shape estimation and intrinsic alignments, both of which

are discussed in Section 1.4.2 below, do not apply to galaxy clustering, while model uncertainties in the relationship between the positions of galaxies relative to their local dark matter distribution are irrelevant to weak lensing. Combining the two types of data can therefore reduce the importance of these sources of systematic error, while increasing the statistical power of the analysis (e.g. Abbott et al. 2018).

Combined weak lensing and galaxy clustering analyses typically also include the cross-correlation of galaxy positions and shapes, known as galaxy-galaxy lensing. As well as probing the evolution of large-scale structure, galaxy-galaxy lensing offers valuable insight into galaxy evolution (e.g. Zacharegkas et al. 2022).

1.4.2. Challenges in weak lensing

Weak lensing analyses are fraught with many challenges and sources of systematic error. Some of the most significant of these are now discussed.

Intrinsic alignments

It is not necessarily safe to assume that all the source galaxies in a weak lensing analysis are randomly oriented prior to any lensing. Tidal forces during galaxy formation can lead to galaxies being aligned to their local large-scale dark matter structure. This leads to a preferential alignment among physically nearby galaxies, which could be misinterpreted as a lensing signal. Additionally, a foreground galaxy could be intrinsically aligned to its local large-scale structure, which lenses a background galaxy, also causing a spurious signal—this time an anti-correlation. Both effects have been measured in real data (Brown et al. 2002; Hirata & Seljak 2004; Joachimi et al. 2011; Mandelbaum et al. 2011; Singh, Mandelbaum & More 2015).

The effect of intrinsic alignments may be mitigated in broadly one of two ways: either by downweighting certain nearby galaxy pairs (Heymans & Heavens 2003; Heymans et al. 2004; Joachimi & Schneider 2008) or by modelling the effect directly (Bridle & King 2007; Schneider & Bridle 2010; Blazek, Vlah & Seljak 2015; Blazek et al. 2019; Fortuna et al. 2021; Samuroff, Mandelbaum & Blazek 2021; Harnois-Déraps, Martinet & Reischke 2022). Since the processes that cause intrinsic alignments are not well understood, such models typically involve several free parameters which must be marginalised over to incorporate the model uncertainty into cosmological parameter constraints (e.g. Joachimi & Bridle 2010; Troxel & Ishak 2015; Amon et al. 2022).

Redshift determination

In an ideal scenario, the redshift of each galaxy in a weak lensing analysis would be determined using spectroscopy: measuring the full spectrum of the galaxy, identifying key

features with known rest-frame wavelengths and comparing to their observed wavelengths to determine the redshift with high confidence. However, the galaxies in a real weak analysis are too numerous and too faint to determine spectroscopic redshifts for all but a small fraction of the sample. Instead, photometric redshifts are used: galaxy flux is measured in a few different bands, and these measurements are used to estimate the redshift of the galaxy. This is a necessary but highly uncertain method, which has the potential to introduce serious biases into cosmological results if not properly treated (Sun et al. 2009; Hearin et al. 2010). The optimal treatment of photometric redshifts and the reduction of potential induced biases are significant areas of active research within weak lensing (D’Isanto & Polsterer 2018; Graham et al. 2018; Bilicki et al. 2018; Pasquet et al. 2019; Amaro et al. 2019; Leistedt et al. 2019; Boucaud et al. 2020; Wright et al. 2020; Schmidt et al. 2020; Schuldt et al. 2021; Henghes et al. 2021; Cordero et al. 2022; Lima et al. 2022; Rau et al. 2022). In practice, much uncertainty remains, and it is therefore still necessary to marginalise over many nuisance parameters describing photometric redshift uncertainty in a real analysis (e.g. Heymans et al. 2021; DES Collaboration et al. 2022).

Shape determination

There are several challenges to determining galaxy shapes sufficiently accurately and precisely to detect a cosmic shear signal, which is an effect of order per cent. One such challenge is the point spread function (PSF) of the telescope, which must be known to a very high precision in order to avoid misinterpreting its effect as a shear signal (Kuijken 1999; Jarvis & Jain 2004; Rhodes et al. 2007; Rowe 2010; Chang et al. 2012; Lu et al. 2017; Eriksen & Hoekstra 2018; Zhang, Mandelbaum & The LSST Dark Energy Science Collaboration 2022). Another is overlapping galaxies on the sky, known as blending, which is an inevitable consequence of the large numbers of galaxies included in weak lensing analyses. Mistakenly identifying two overlapping galaxies as a single galaxy can cause biases, but so too can naively removing galaxies identified as overlapping, since this would create a selection bias against higher density regions (Dawson et al. 2016; Hoekstra, Kannawadi & Kitching 2021; Gaztanaga et al. 2021; Sanchez et al. 2021; Melchior et al. 2021; Nourbakhsh et al. 2022). A third class of challenges in shape determination is detector systematics, such as charge transfer inefficiency (Rhodes et al. 2010) and the brighter-fatter effect (Walter 2015; Gilbertson, Nomerotski & Takacs 2017; Coulton et al. 2018; Rowlands, Midwinter & Warner 2018; Freudenburg et al. 2020).

Modelling of non-linear scales

Regardless of how accurately and precisely observational measurements can be made, final cosmological results are still limited by the ability to model all physical effects reliably. A particular challenge in this regard lies in the modelling of small physical scales. The growth of structure on such scales and the effect of astrophysical feedback processes are

poorly understood and difficult to model. There is disagreement between different models, and between model predictions and simulations (Casarini et al. 2011; Schneider et al. 2016; Giblin et al. 2019; Bose, Koyama & Winther 2019; Bose et al. 2020). This is a highly active area of ongoing study, and is likely to limit the scales at which future weak lensing data can be analysed reliably (Huterer & Takada 2005; Jing et al. 2006; van Daalen et al. 2011; Semboloni et al. 2011; Semboloni, Hoekstra & Schaye 2013; Zentner et al. 2013; Mohammed et al. 2014; Eifler et al. 2015; MacCrann et al. 2017; Copeland, Taylor & Hall 2018; Huang, Addison & Bennett 2019; Huang et al. 2021; Martinelli et al. 2021). Choosing such scale cuts carefully can isolate well-understood scales while extracting maximal cosmological information (Kitching & Taylor 2011; Taylor, Bernardeau & Kitching 2018; Deshpande, Taylor & Kitching 2020; Taylor, Bernardeau & Huff 2021; Taylor et al. 2021).

Cosmic variance

Cosmic variance describes the principle that the value of any cosmological observable measured from Earth is just one sample from the distribution of this observable over the entire Universe, which of course extends far beyond the most distant galaxies in a weak lensing survey (Kamionkowski & Loeb 1997; Wiltshire 2007; Driver & Robotham 2010; Moster et al. 2011). In practice, what this means is that even given perfect knowledge of both theory and experiment, the value of a cosmological observable cannot be predicted exactly; it is instead given by a probability distribution. In some ideal cases, we may know the form of this distribution, but more generally it is necessary to make approximations. Inverting the argument above implies that, even with perfect knowledge of an experiment and no uncertainty of any kind on the observed data, we will necessarily still have non-zero error bars on the inferred cosmological parameters (e.g. Martel, Shapiro & Weinberg 1998; Taylor & Kitching 2010).

A consequence of cosmic variance is that it demands a statistical treatment of all observables in cosmology. As the precision achieved by cosmological experiments increases, so too must the level of understanding of the statistical properties of all aspects of the data. Upcoming weak lensing experiments such as *Euclid* (see Section 1.4.4) will offer an unprecedented level of precision, due to the extremely large numbers of observed galaxies. This level of precision demands an equally unprecedented level of refinement in statistical modelling, which is the main focus of this thesis. The mathematical concepts involved in this statistical modelling of cosmology, and specifically weak lensing, will be introduced in Chapter 2.

1.4.3. Observational history of weak lensing

The first successful observations of weak lensing by large scale structure were made in 2000 by four separate teams. Van Waerbeke et al. (2000) used the Canada France Hawaii

Telescope (CFHT) to survey 191 000 galaxies over an area of 6300 arcmin², detecting cosmic shear on scales from 0.5 to 3.5 arcmin. Wittman et al. (2000) claimed a detection on scales up to 30 arcmin, from 145 000 galaxies imaged using the Big Throughput Camera (Wittman et al. 1998) on the 4 m Blanco telescope at the Cerro Tololo Inter-American Observatory. Detections were also made by Bacon, Refregier & Ellis (2000) using the 4.2 m William Herschel Telescope (Boksenberg 1985) over a total area of 1792 arcmin², and by Kaiser, Wilson & Luppino (2000) using the CFHT with around 120 000 galaxies over an area of 5400 arcmin².

More surveys followed in the early 2000s. The first space-based weak lensing survey was carried out using the Hubble Space Telescope (HST), surveying 4000 galaxies in the 168 arcmin² HST survey strip using the Wide Field Planetary Camera 2 (Griffiths 1990), to place constraints on S_8 (Equation 1.46) with a one-third error bar at 68% posterior probability (Rhodes, Refregier & Groth 2001). This was improved upon with the Red-Sequence Cluster Survey (Gladders & Yee 2001), which covered 53 deg² using both the CFHT and Cerro Tololo Inter-American Observatory 4 m telescope and included 1 773 543 galaxies, to place constraints on S_8 with a 27% error at 95% posterior probability (Hoekstra et al. 2002; Hoekstra, Yee & Gladders 2002). The COMBO-17 survey covered 1.25 deg² using the La Silla 2.2 m telescope in Chile (Wolf et al. 2001), including 83 514 galaxies, and further improved constraints on S_8 to 12.5% at 68% posterior probability (Brown et al. 2003). Further surveys followed: 2.1 deg² using the Suprime-Cam (Miyazaki et al. 2002) at the Subaru Telescope (Hamana et al. 2003); the Cerro Tololo Inter-American Observatory Survey of 2 million galaxies covering 75 deg² (Jarvis et al. 2003); the VIRMOS-Descart survey using the CFHT (Van Waerbeke, Mellier & Hoekstra 2005); and the HST GEMS Survey using the Advanced Camera for Surveys, including 71 233 galaxies over 795 arcmin² in the *Chandra* Deep Field South with an extremely high galaxy number density of 60 / arcmin², compared to a more typical value of ~ 30 / arcmin² (Heymans et al. 2005).

The CFHT Lensing Survey (CFHTLenS; Heymans et al. 2012) was the first to have sufficient numbers of galaxies across a range of redshifts to place constraints on the expansion of the Universe and its acceleration, via the dimensionless Hubble parameter h and the dark energy equation of state parameter w . CFHTLenS was an optical imaging survey based on 154 deg² of data taken with the CFHT between 2003 and 2009 as part of the CFHT Legacy Survey. It included around 6 million galaxies with shape and photometric redshift estimates, across a redshift range from $z = 0.5$ to 1.3. The survey produced a number of constraints on cosmology (Kilbinger et al. 2013; Benjamin et al. 2013; Heymans et al. 2013; Van Waerbeke et al. 2013; Kitching et al. 2014; Fu et al. 2014) including values of $h = 0.78 \pm 0.12$ and $w_0 = -1.10 \pm 0.15$ (68% credible interval) when combined with WMAP data. There have also been several reanalyses combining CFHTLenS data with other external data (e.g. More et al. 2015; Blake et al. 2016; Joudaki et al. 2017).

Three large weak lensing surveys are currently ongoing. This generation are known as

Stage III surveys, based on the definitions set out in the Report of the Dark Energy Task Force (Albrecht et al. 2006) regarding their ability to constrain dark energy. These are the Hyper Suprime-Cam Subaru Strategic Program Survey, the Kilo-Degree Survey, and the Dark Energy Survey.

The Hyper Suprime-Cam Subaru Strategic Program (HSC SSP; Aihara et al. 2018) is a 1400 deg^2 optical imaging survey with the HSC (Miyazaki et al. 2012) on the 8.2 m Subaru Telescope in Hawaii. To date an area of 670 deg^2 has been observed and released publicly (Aihara et al. 2022), but the only cosmological analysis so far is of the Year 1 data, comprising 9 million galaxies over 137 deg^2 with a redshift range of $z = 0.3\text{--}1.5$. This obtained a constraint of $S_8 = 0.780^{+0.030}_{-0.033}$ at 68% posterior probability (Hikage et al. 2019). No meaningful constraints on dark energy were possible with the first year data alone, but should be possible with the ~ 90 million galaxies expected to be observed in the full survey.

The Kilo-Degree Survey (KiDS; de Jong et al. 2013) is a 1350 deg^2 optical imaging survey, carried out with the OmegaCAM (Kuijken 2011) on the VLT Survey Telescope at the Paranal Observatory in Chile. To date, an area of 1006 deg^2 has been observed and released publicly (KiDS-1000; Kuijken et al. 2019). The main cosmological analysis from this data set included 21 million galaxies over a redshift range of $z = 0.1\text{--}1.2$ (Giblin et al. 2021), and obtained a constraint of $S_8 = 0.766^{+0.020}_{-0.014}$. As mentioned in Section 1.2, this value is discrepant with that from *Planck* CMB data (Planck Collaboration et al. 2020b) at $\sim 2\text{--}3\sigma$ (Heymans et al. 2021).

The Dark Energy Survey (DES; Dark Energy Survey Collaboration 2005) is a six-year optical imaging survey on the Dark Energy Camera (Flaugh et al. 2015) at the Blanco 4 m telescope at the Cerro Tololo Inter-American Observatory in Chile. 300 million galaxies were observed between 2013 and 2019, but to date only 100 million have been included in published cosmological analyses (DES Collaboration et al. 2022). The survey covers 5000 deg^2 and includes galaxies out to redshift $z = 2$. The Year 3 cosmological analysis reports results of $S_8 = 0.775^{+0.026}_{-0.024}$ and $w = -0.98^{+0.32}_{-0.20}$ from DES data alone. Combined with *Planck* CMB data, plus baryon acoustic oscillations, redshift-space distortions and Type Ia supernovae, they obtain $S_8 = 0.812 \pm 0.08$; $w = -1.031^{+0.030}_{-0.027}$; $h = 0.687^{+0.006}_{-0.007}$. Some parameter constraints from the Year 3 analysis are shown in Figure 1.7. To obtain these results, such methodological advances were necessary that the collaboration published 29 methods papers for the Year 3 analysis alone (DES Collaboration et al. 2022, and references therein). The future Year 6 analysis will require still further refinement to reliably extract the maximum information from all 300 million sources.

1.4.4. *Euclid* satellite

Euclid (Laureijs et al. 2011) is an upcoming European Space Agency (ESA) satellite mission. It was born from a combination of two satellite proposals to ESA: the Dark Universe Ex-

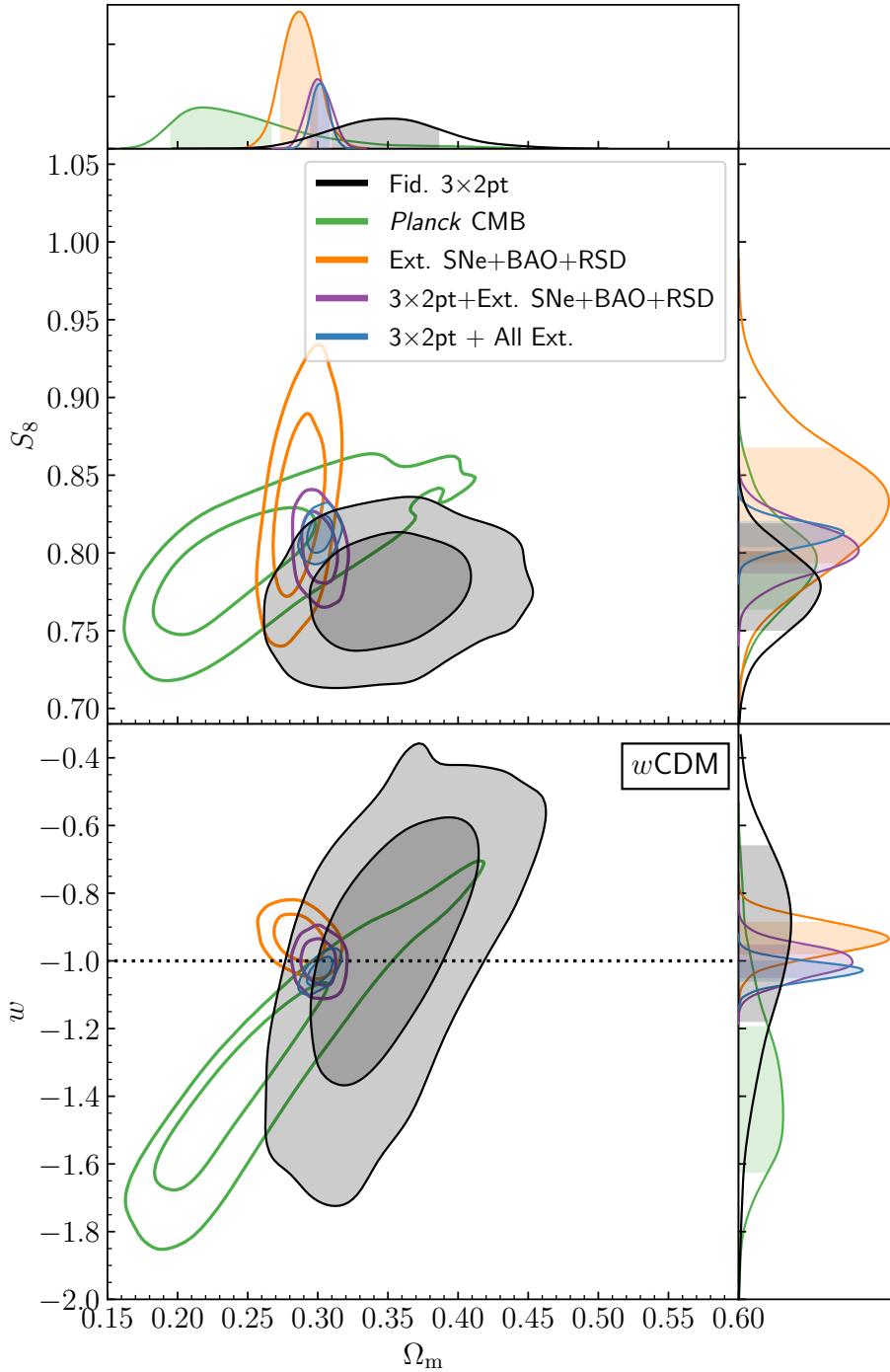


Figure 1.7. Cosmological constraints from the Dark Energy Survey Year 3 analysis (DES Collaboration et al. 2022). Each panel shows the 68% and 95% credible regions of joint constraints on two parameters marginalised over all other parameters. The grey contours show the constraints from DES data alone, while the blue constraints include *Planck* CMB data, baryon acoustic oscillations, redshift-space distortions and Type Ia supernovae. Other colours show other combinations of data.

plorer (DUNE; Refregier & The DUNE Collaboration 2009), a weak lensing mission, and the Spectroscopic All Sky Cosmic Explorer (SPACE; Cimatti et al. 2009), a spectroscopic galaxy clustering mission to measure baryon acoustic oscillation and redshift-space distortions. As such, *Euclid* is equipped with two complementary instruments: an optical imager—VIS—and a near-infrared spectrometer & photometer—NISP. *Euclid* is scheduled to be launched in early 2023 from the ESA spaceport in French Guiana, into an orbit around the L_2 Sun–Earth Lagrange point.

Over a nominal lifetime of six years, *Euclid* will survey $15\,000 \text{ deg}^2$ of the extragalactic sky (the Euclid Wide Survey; Euclid Collaboration: Scaramella et al. 2022) out to a magnitude of 26.2 with VIS and 24.5 with NISP. It is expected to survey around 10 billion galaxies in total, with around 1.5 billion having sufficiently precise shape and photometric redshift estimates for use in the weak lensing analysis. These source galaxies will have redshifts out to $z \sim 2$, with the majority having $z \sim 1$. Around 23 million galaxies will have precise spectroscopic redshift estimates made using NISP. There will also be three deep fields, together covering 40 deg^2 to form the Euclid Deep Survey, which will reach about two magnitudes deeper than the wide survey.

The *Euclid* cosmological analysis will use weak lensing and spectroscopic galaxy clustering, plus photometric galaxy clustering and its cross-correlation with weak lensing (Euclid Collaboration: Blanchard et al. 2020). Using these probes, it will determine the late-time evolution of the large-scale matter distribution in the Universe, and measure baryon acoustic oscillations and redshift-space distortions. The main scientific goal is to constrain dark energy, constraints on which are expected to improve by over an order of magnitude relative to current Stage III surveys (Harrison et al. 2016). *Euclid* will also be able to tightly constrain other parameters of the $w\text{CDM}$ model, as shown in Figure 1.8, as well as theories of modified gravity.

The preparation and initial data analysis is conducted by the Euclid Consortium, which currently contains over 1400 people in 17 countries globally. After the initial analysis, data will be released to the public, where it is expected to have a big impact on legacy science. Potential non-cosmology uses include studies of galactic physics, and stellar physics within the Milky Way.

1.4.5. Other future surveys

There are at least three other Stage IV weak lensing surveys planned alongside *Euclid*. The first is the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; formerly the Large Synoptic Survey Telescope; Ivezić et al. 2019). Rubin is a brand new ground-based observatory in Chile, which will conduct an optical imaging survey of around $18\,000 \text{ deg}^2$ of the southern sky. It has similar goals to *Euclid* of constraining dark energy and structure growth, with some complementary characteristics; for example, *Euclid* will have higher

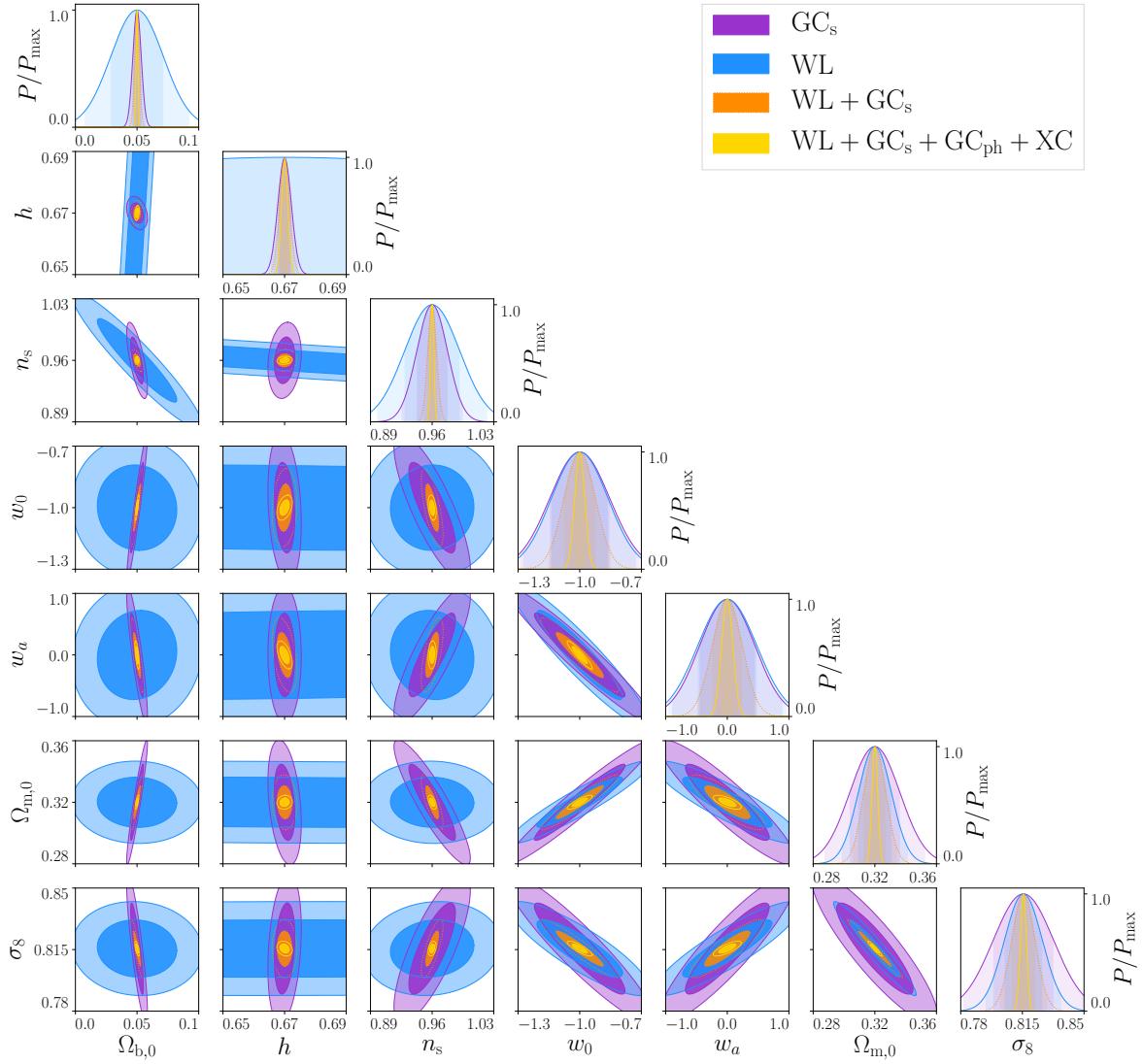


Figure 1.8. Forecasted constraints for w CDM parameters from future *Euclid* data, combining spectroscopic galaxy clustering (GC_s), weak lensing (WL), photometric galaxy clustering (GC_{ph}) and cross-correlations (XC). Taken from Euclid Collaboration: Blanchard et al. (2020).

angular resolution, but Rubin will be able to detect fainter objects. This means that there is great potential from combining data from the two experiments (Rhodes et al. 2017; Guy et al. 2022). The survey is scheduled to last for ten years, starting in 2023.

Euclid is expected to be joined at the L_2 point by the Nancy Grace Roman Space Telescope (formerly the Wide-Field Infrared Survey Telescope; Spergel et al. 2015), a NASA satellite currently scheduled to launch by May 2027. It will carry two instruments: a wide-field optical and near-infrared camera used primarily for a galaxy survey, alongside a coronograph for directly imaging exoplanets. Like *Euclid* and Rubin, the main cosmological aim of Roman is to constrain dark energy. To do so, it will survey around 2000 deg^2 of the extragalactic sky over a period of around 24 months. It will be sensitive to magnitudes up to 28.5, so should be able to detect fainter galaxies than *Euclid*, with a superior angular resolution due to its larger telescope, to compensate for the smaller survey area.

Finally, the Square Kilometre Array (SKA; Square Kilometre Array Cosmology Science Working Group et al. 2020) is expected to be able to perform weak lensing analyses using radio observations. The SKA is a radio array currently under construction in Australia and South Africa, expected to begin observations in some capacity in the next decade. Use of the SKA for weak lensing has several key advantages compared to optical surveys such as *Euclid* (Harrison et al. 2016): it will probe a broader redshift distribution, potentially out to $z \sim 6$; there is less PSF contamination for radio observations; intrinsic alignments may be able to be directly measured using radio polarisation (Brown & Battye 2011; Whittaker, Brown & Battye 2015); and it has the ability to access more of the sky, since the Galaxy is effectively transparent to radio interferometers. However, the greatest benefits are gained by cross-correlating radio and optical data, since this not only unlocks additional constraining power but also significantly reduces the impact of systematic errors, which are largely different for radio and optical surveys (Camera et al. 2017).

Chapter 2

Cosmological estimators and likelihoods

This chapter introduces some of the mathematical concepts involved in the later chapters in this thesis. It begins with an introduction to cosmological fields on the sphere in Section 2.1, before focusing specifically on weak lensing fields and two-point statistics in Section 2.2. Section 2.3 introduces the pseudo- C_ℓ method, which is central to much of the work presented in this thesis, before the concepts of Bayesian inference and likelihoods are described in Section 2.4.

This chapter has been compiled from Heavens (2003), Chon et al. (2004), Brown, Castro & Taylor (2005), Kilbinger et al. (2017), Chisari et al. (2019), and Fang et al. (2020).

2.1. Cosmological fields on the sphere

The sky as viewed from the Earth, or indeed anywhere, may be modelled as a sphere (though one can only see up to half the sphere from a fixed point on the surface of the Earth at any given instant). As a result, astronomical observations may be described as fields on the sphere. In the context of cosmology, cosmic variance (introduced in Chapter 1) means that cosmological fields are modelled as random fields on the sphere. The cosmological principle requires that these fields are statistically isotropic and translation invariant, meaning that information is only contained in the dependence on angular scales and not in orientation or position. Because of this property, it is often most convenient to work with spherical harmonics. For a general scalar field ψ , the transforms between the field in angular coordinates $\psi(\theta, \phi)$ and in spherical harmonics $\psi_{\ell m}$ are

$$\psi_{\ell m} = \int d\theta \int d\phi \psi(\theta, \phi) Y_{\ell m}^*(\theta, \phi); \quad (2.1)$$

$$\psi(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} \psi_{\ell m} Y_{\ell m}(\theta, \phi), \quad (2.2)$$

where $Y_{\ell m}(\theta, \phi)$ are the spherical harmonics, and $*$ denotes complex conjugation. Equations (2.1)–(2.2) are the spin-0 spherical harmonic transforms (spin will be discussed in Section 2.1.1).

Statistical isotropy implies that only the degree ℓ and not the order m carries cosmological information. The field is then mostly characterised by the angular power spectrum C_ℓ , defined as the expectation value of the product of spherical harmonic coefficients,

$$\langle \psi_{\ell m} \psi_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^\psi, \quad (2.3)$$

where δ is the Kronecker delta function. If the field ψ follows Gaussian statistics, the information content of the field is entirely contained within this power spectrum. If the field is non-Gaussian, higher-order moments are required to fully describe it.

An equivalent, alternative statistic to the angular power spectrum is the angular two-point correlation function $\xi(\theta)$, defined as

$$\langle \psi(\Omega) \psi(\Omega') \rangle = \xi^\psi(|\Omega' - \Omega|), \quad (2.4)$$

where Ω and Ω' are sky coordinates. The correlation function may be obtained from the power spectrum as

$$\xi(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_\ell d_{00}^\ell(\theta), \quad (2.5)$$

where $d_{m'm}^\ell$ is a Wigner small- d symbol.

Two correlated fields α and β may be described by their cross-power spectrum $C_\ell^{\alpha\beta}$,

$$\langle \alpha_{\ell m} \beta_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{\alpha\beta}, \quad (2.6)$$

or by their cross-correlation function, defined analogously to Equation (2.4). Cross-power spectra and cross-correlation functions are symmetric such that

$$C_\ell^{\alpha\beta} = C_\ell^{\beta\alpha}; \quad (2.7)$$

$$\xi^{\alpha\beta}(\theta) = \xi^{\beta\alpha}(\theta). \quad (2.8)$$

2.1.1. Spin

Some cosmological fields are scalar, in that they are described by a single value at each point on the sky. Examples include the CMB temperature anisotropies, or the number density of galaxies. These fields are spin-0, and obey the above equations.

Some cosmological fields are spin-2, such as the CMB polarisation, and weak lensing shear. These fields are described by two values at each point, such that they can be modelled as complex. A general spin-2 field γ can be decomposed into two components γ_1 and γ_2 ,

$$\gamma = \gamma_1 + i\gamma_2, \quad (2.9)$$

or into a magnitude $|\gamma|$ and phase ϕ ,

$$\gamma = |\gamma| e^{2i\phi}. \quad (2.10)$$

The factor of 2 in Equation (2.10) captures the spin-2 nature of γ . It ensures that the field is invariant under

$$\phi \rightarrow \phi + \pi. \quad (2.11)$$

The equivalent of Equations (2.1)–(2.2) for spin-2 fields are the spin-2 spherical harmonic transforms,

$${}_2\gamma_{\ell m} = \int d\theta \int d\phi \gamma(\theta, \phi) {}_2Y_{\ell m}^*(\theta, \phi); \quad (2.12)$$

$$-{}2\gamma_{\ell m} = \int d\theta \int d\phi \gamma^*(\theta, \phi) -{}2Y_{\ell m}^*(\theta, \phi); \quad (2.13)$$

$$\gamma(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} {}_2\gamma_{\ell m} {}_2Y_{\ell m}(\theta, \phi); \quad (2.14)$$

$$\gamma^*(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} -{}2\gamma_{\ell m} -{}2Y_{\ell m}(\theta, \phi), \quad (2.15)$$

where ${}_s Y_{\ell m}(\theta, \phi)$ are the spin-weighted spherical harmonics.

An alternative decomposition of a spin-2 field that is often more closely related to the physical origins of the field (such as in both CMB polarisation and weak lensing shear) is the decomposition into *E*-mode and *B*-mode components, as

$$E_{\ell m} = \frac{1}{2} [{}_2\gamma_{\ell m} + -{}2\gamma_{\ell m}]; \quad (2.16)$$

$$B_{\ell m} = \frac{-i}{2} [{}_2\gamma_{\ell m} - -{}2\gamma_{\ell m}]. \quad (2.17)$$

It is then possible to transform directly between the spin-2 field and its *E*- and *B*-mode harmonic coefficients as

$$E_{\ell m} = \frac{1}{2} \int d\theta \int d\phi [\gamma(\theta, \phi) {}_2Y_{\ell m}^*(\theta, \phi) + \gamma^*(\theta, \phi) -{}2Y_{\ell m}^*(\theta, \phi)]; \quad (2.18)$$

$$B_{\ell m} = \frac{-i}{2} \int d\theta \int d\phi [\gamma(\theta, \phi) {}_2Y_{\ell m}^*(\theta, \phi) - \gamma^*(\theta, \phi) -{}2Y_{\ell m}^*(\theta, \phi)]; \quad (2.19)$$

$$\gamma(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} [E_{\ell m} + iB_{\ell m}] {}_2Y_{\ell m}(\theta, \phi); \quad (2.20)$$

$$\gamma^*(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} [E_{\ell m} - iB_{\ell m}] -{}2Y_{\ell m}(\theta, \phi). \quad (2.21)$$

We may then define the *E*-mode and *B*-mode power spectra, C_{ℓ}^{EE} and C_{ℓ}^{BB} , as

$$\langle E_{\ell m} E_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{EE}; \quad (2.22)$$

$$\langle B_{\ell m} B_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{BB}, \quad (2.23)$$

as well as the E - B cross-power spectrum C_ℓ^{EB} ,

$$\langle E_{\ell m} B_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{EB}. \quad (2.24)$$

These may be used to obtain the spin-2 ξ^+ and ξ^- correlation functions,

$$\xi^\pm(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} [C_\ell^{EE} \pm C_\ell^{BB}] d_{\pm 22}^\ell(\theta). \quad (2.25)$$

There may also be cross-power spectra between spin-0 and spin-2 fields,

$$\langle \psi_{\ell m} E_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{\phi E}; \quad (2.26)$$

$$\langle \psi_{\ell m} B_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{\phi B}, \quad (2.27)$$

and a corresponding cross-correlation function ξ^\times ,

$$\xi^\times(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} [C_\ell^{\phi E} - i C_\ell^{\phi B}] d_{20}^\ell(\theta). \quad (2.28)$$

2.2. Weak lensing fields and two-point statistics

2.2.1. Shear field

The main weak lensing target observable field is the shear field, γ . In the weak lensing regime, the shear is additive to the galaxy ellipticity, such that the observed ellipticity of a galaxy ϵ is given by

$$\epsilon = \epsilon_{\text{int}} + \gamma, \quad (2.29)$$

where ϵ_{int} is the intrinsic galaxy ellipticity. Taking the two-point correlation of the ellipticity gives

$$\begin{aligned} \langle \epsilon \epsilon^* \rangle &= \langle (\epsilon_{\text{int}} + \gamma) (\epsilon_{\text{int}} + \gamma)^* \rangle \\ &= \langle \epsilon_{\text{int}} \epsilon_{\text{int}}^* \rangle + \langle \gamma \gamma^* \rangle + \langle \epsilon_{\text{int}} \gamma^* \rangle + \langle \gamma \epsilon_{\text{int}}^* \rangle. \end{aligned} \quad (2.30)$$

The first term in Equation (2.30) corresponds to shape noise (see Section 2.2.4), while the third and fourth correspond to intrinsic alignments, discussed in Chapter 1. In practice all of these terms contribute to the observed two-point correlation of galaxy ellipticity, but in an idealised scenario with no intrinsic alignments and where an infinite number of galaxies

are averaged over, we have

$$\langle \epsilon \epsilon^* \rangle = \langle \gamma \gamma^* \rangle. \quad (2.31)$$

The shear field is spin-2, so follows the spin-2 spherical harmonic transforms described in Section 2.1.1 above. It has the important property that the $s = +2$ and $s = -2$ spherical harmonic coefficients are identical,

$${}^2\gamma_{\ell m} = {}^{-2}\gamma_{\ell m} \equiv \gamma_{\ell m}. \quad (2.32)$$

The shear field at a particular redshift z and a point on the sky Ω , $\gamma(z, \Omega)$, is related to the lensing potential field $\phi(z, \Omega)$ via their respective harmonic coefficients, as

$$\gamma_{\ell m}(z, \Omega) = \frac{1}{2} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \phi_{\ell m}(z, \Omega). \quad (2.33)$$

The lensing potential is a spin-0 field, which is given by a projection of the three-dimensional gravitational potential $\Phi(z, \Omega)$, as

$$\phi(z, \Omega) = \frac{2}{c^2} \int_0^z dz' \left[\frac{S_K(r) - S_K(r')}{S_K(r) S_K(r')} \right] \Phi(z', \Omega), \quad (2.34)$$

where $r = \chi(z)$, and $S_K(\chi)$ is defined in Equation (1.12).

The gravitational potential is related to the matter overdensity field $\delta(z, \Omega)$ by the Poisson equation,

$$\nabla^2 \Phi(z, \Omega) = \frac{3\Omega_m H_0^2}{2a(t)} \delta(z, \Omega), \quad (2.35)$$

where the matter density parameter Ω_m , the Hubble constant H_0 and the scale factor $a(t)$ are all defined in Chapter 1. The matter overdensity field $\delta(z, \Omega)$ is defined in terms of the density field $\rho(z, \Omega)$ as

$$\delta(z, \Omega) = \frac{\rho(z, \Omega) - \bar{\rho}}{\bar{\rho}}, \quad (2.36)$$

where $\bar{\rho}$ is the mean density.

2.2.2. Galaxy number overdensity field

For galaxy clustering it is simplest to consider the galaxy number overdensity field $n(z, \Omega)$, which is a spin-0 field defined analogously to Equation (2.36) in terms of the galaxy number density field $N(z, \Omega)$ as

$$n(z, \Omega) = \frac{N(z, \Omega) - \bar{N}}{\bar{N}}. \quad (2.37)$$

2.2.3. 3×2 pt power spectra

For a general set of galaxy number overdensity fields $n(z_i)$ and shear fields $\gamma(z_i)$, we may form three types of two-point correlations: galaxy–galaxy, shear–shear, and galaxy–shear. This common set of three two-point correlations is known as 3×2 pt. The six 3×2 pt power spectra are defined in terms of the spherical harmonic coefficients of the spin-0 number overdensity fields $n_{\ell m}(z_i)$ and of the E - and B -mode components of the spin-2 shear fields $E_{\ell m}(z_i)$ and $B_{\ell m}(z_i)$ as (henceforth dropping the angular coordinate Ω for clarity)

$$\langle n_{\ell m}(z_1) n_{\ell' m'}^*(z_2) \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{n(z_1)n(z_2)}; \quad (2.38)$$

$$\langle E_{\ell m}(z_1) E_{\ell' m'}^*(z_2) \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{E(z_1)E(z_2)}; \quad (2.39)$$

$$\langle B_{\ell m}(z_1) B_{\ell' m'}^*(z_2) \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{B(z_1)B(z_2)}; \quad (2.40)$$

$$\langle n_{\ell m}(z_1) E_{\ell' m'}^*(z_2) \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{n(z_1)E(z_2)}; \quad (2.41)$$

$$\langle n_{\ell m}(z_1) B_{\ell' m'}^*(z_2) \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{n(z_1)B(z_2)}; \quad (2.42)$$

$$\langle E_{\ell m}(z_1) B_{\ell' m'}^*(z_2) \rangle = \delta_{\ell \ell'} \delta_{mm'} C_\ell^{E(z_1)B(z_2)}. \quad (2.43)$$

A consequence of Equation (2.32) is that the B -mode component of the shear field vanishes, such that

$$C_\ell^{B(z_1)B(z_2)} = 0; \quad (2.44)$$

$$C_\ell^{n(z_1)B(z_2)} = 0; \quad (2.45)$$

$$C_\ell^{E(z_1)B(z_2)} = 0. \quad (2.46)$$

However, the BB power spectrum will contain a noise contribution when $z_1 = z_2$ (see Section 2.2.4 below). B -modes in the observed ellipticity field can also be introduced by intrinsic alignments and other systematic effects. In practice, B -mode power spectra are often tested for consistency with zero as a check of systematics (e.g. Asgari et al. 2019).

The remaining three power spectra $C_\ell^{X(z_1)Y(z_2)}$, where each of X and Y may be n or E , are given by a projection over the matter distribution,

$$C_\ell^{X(z_1)Y(z_2)} = \frac{2}{\pi} \int_0^\infty \frac{dk}{k} k^3 \int_0^{z_1} dz \int_0^{z_2} dz' P_m(k, z, z') j_\ell(k\chi(z)) j_\ell(k\chi(z')) w_\ell^X(k, z) w_\ell^Y(k, z'), \quad (2.47)$$

with respective weight functions for galaxy number overdensity and shear given by

$$w_\ell^N(k, z) = n(z) b(z); \quad (2.48)$$

$$w_\ell^E(k, z) = \frac{3H_0^2\Omega_m}{2k^2} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \frac{(1+z)}{cH(z)} \int_z^\infty dz' n(z') \frac{\chi(z') - \chi(z)}{\chi(z') \chi(z)}, \quad (2.49)$$

where additional effects such as redshift-space distortions, magnification and intrinsic alignments have been neglected. j_ℓ are the spherical Bessel functions of order ℓ , $n(z)$ is the normalised redshift distribution of galaxies in the survey, $b(z)$ is the linear galaxy bias such that $n(z) = b(z)\delta(z)$, and the matter distribution is described by the matter power spectrum $P_m(k, z, z')$, defined as

$$\langle \tilde{\delta}(\mathbf{k}, z) \tilde{\delta}^*(\mathbf{k}', z') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_m(|\mathbf{k}|, z, z'), \quad (2.50)$$

where $\tilde{\delta}(\mathbf{k}, z)$ is the Fourier transform of the matter overdensity field at redshift z , and δ_D is the Dirac delta function.

It follows from Equation (2.47) that measuring a set of 3×2 pt power spectra over a range of redshifts directly probes the evolution of the matter distribution. This offers a wealth of information about recent structure growth and the expansion of the Universe, the details of which depend closely on the properties of dark energy and dark matter, as described in Chapter 1. This is why weak lensing, and especially the combination of weak lensing and galaxy clustering, are such valuable cosmological probes.

2.2.4. Shape and shot noise

The ability to trace the underlying matter distribution by the shapes and positions of galaxies is limited by the finite number of galaxies in a survey. This introduces a noise term, resulting from the correlation of each galaxy with itself, which is suppressed as more galaxies are included. For galaxy clustering this noise term depends only on the number density of galaxies in a particular redshift bin, $N_s(z)$, and is known as shot noise. For cosmic shear, there is an additional dependence on the dispersion of intrinsic shapes σ_ϵ , which is defined per component such that it contributes equally to the E - and B -mode power spectra. This is known as shape noise. The noise contributions to the power spectra, N_ℓ , are given by

$$N_\ell^{n(z_1)n(z_2)} = \delta_{z_1 z_2} \frac{1}{N_s(z_1)}; \quad (2.51)$$

$$N_\ell^{E(z_1)E(z_2)} = N_\ell^{B(z_1)B(z_2)} = \delta_{z_1 z_2} \frac{\sigma_\epsilon^2}{N_s(z_1)}; \quad (2.52)$$

$$N_\ell^{n(z_1)E(z_2)} = N_\ell^{n(z_1)B(z_2)} = N_\ell^{E(z_1)B(z_2)} = 0. \quad (2.53)$$

Since the noise power spectra are inversely proportional to the number density of galaxies, this is a motivation for Stage IV weak lensing surveys to aim to detect large numbers of galaxies.

2.3. Pseudo- C_ℓ method

For full-sky observations, the power spectrum is the underlying covariance of the spherical harmonic coefficients, as was shown for the general case in Equation (2.6). (Covariance is defined in Equation (2.69) below.) Therefore, the power spectrum may be estimated simply using the sample covariance of those coefficients,

$$\hat{C}_\ell^{\alpha\beta} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{+\ell} \alpha_{\ell m} \beta_{\ell m}^*. \quad (2.54)$$

However, in practice full-sky observations are never truly possible. From the ground, some parts of the sky are never visible, while even in space much of the sky is obscured by the Galaxy and other nearby objects (for example, *Euclid* will only observe around one third of the sky). The estimator in Equation (2.54) does not return the underlying power spectrum if the sky is cut, because the full sphere is needed in order to cleanly decompose into spherical harmonics (see Chapter 3 for more details). Therefore, another approach to estimating power spectra is needed. One such approach is the pseudo- C_ℓ method.

The pseudo- C_ℓ method as presented here was introduced in Hivon et al. (2002) and extended to correlated spin-0 and spin-2 fields in Kogut et al. (2003) and Brown, Castro & Taylor (2005). It has been further modified to include extensions such as contaminant de-projection and E/B purification (see Alonso, Sanchez & Slosar 2019 and references therein). It has been used for CMB (Kogut et al. 2003), galaxy clustering (Camacho et al. 2019), and weak lensing (Hikage et al. 2019) analyses, and will be used for the analysis of future *Euclid* data (Loureiro et al. 2021).

When the estimator in Equation (2.54) is applied to full-sky data, its expected value is equal to the underlying power spectrum,

$$\langle \hat{C}_\ell^{\alpha\beta} \rangle = C_\ell^{\alpha\beta}, \quad (2.55)$$

where $C_\ell^{\alpha\beta}$ may include a noise contribution. However, when it is applied to cut-sky data, denoted here as $\tilde{C}_\ell^{\alpha\beta}$, its expectation value is a linear combination of the power spectrum at different multipoles ℓ ,

$$\langle \tilde{C}_\ell^{\alpha\beta} \rangle = \sum_{\ell'=0}^{\infty} M_{\ell\ell'} C_{\ell'}^{\alpha\beta}, \quad (2.56)$$

where $M_{\ell\ell'}$ are the elements of the mixing matrix \mathbf{M} , which can also mix E - and B -mode

components. Full expressions for the elements of the mixing matrix are given in Brown, Castro & Taylor (2005).

In principle, Equation (2.56) may be inverted to obtain an unbiased estimate of the underlying power spectrum,

$$\left\langle \sum_{\ell'=0}^{\infty} M_{\ell\ell'}^{-1} \tilde{C}_{\ell'}^{\alpha\beta} \right\rangle = C_{\ell}^{\alpha\beta}. \quad (2.57)$$

However, this step is unnecessary and can introduce problems, and is not being used for *Euclid* (Loureiro et al. 2021).

As discussed in Chapter 1, the unprecedented precision offered by *Euclid* demands an equally unprecedented understanding of all aspects of data analysis. This includes the statistical properties of estimators. While the pseudo- C_{ℓ} estimator has been used in previous analyses (e.g. Kogut et al. 2003; Hikage et al. 2019; Camacho et al. 2019; Loureiro et al. 2021), a detailed understanding of its statistical properties is lacking, prior to the work presented in this thesis. This motivates much of the work presented here: Chapters 3, 4, and 5 are dedicated to studying the statistical properties of pseudo- C_{ℓ} estimates.

2.4. Bayesian inference and likelihoods

As a result of cosmic variance, as discussed in Chapter 1, all observable data in cosmology are probabilistic and cannot be predicted exactly by theory. The typical approach to dealing with this property centres on Bayes' theorem, which relates the probability of a given model M (or equivalently, a given set of parameters of a model) to some observed data D ,

$$P(M|D) = \frac{P(D|M) P(M)}{P(D)}. \quad (2.58)$$

$P(M|D)$ in Equation (2.58) is the posterior probability, which is what we want to know: how likely is a particular model, or a particular set of parameter values for a given model. $P(M)$ is the prior, containing the prior knowledge about the probability of M before the observed data D is taken into account. This can incorporate constraints from previous data, or it can be chosen to be uninformative. $P(D)$ is known as the Bayesian evidence; it often acts as simply a normalising factor such that $P(M|D)$ integrates to unity over all values of M , but can also be used to assess the overall fit of the data to a model over all parameter values, in order to compare between models. The remaining term, $P(D|M)$, is the likelihood. This is where the details of the probabilistic nature of the data enter the inference process. Finding the correct likelihood function, or at least a suitable choice, is a challenge, and Chapters 3 and 4 focus on finding a suitable likelihood function for pseudo- C_{ℓ} estimates.

For a given choice of likelihood function, a posterior distribution is obtained by evaluating the likelihood at all values of M (which is often the parameters of a given model) included in the prior, multiplying by the prior itself, and normalising. In practice, the way that the likelihood is evaluated everywhere could be through parameter grid searches, which are used throughout this work, or by random sampling techniques such as Markov Chain Monte Carlo, which are usually necessary in the high-dimensional cosmological analyses of real data.

2.4.1. Credible regions and sigma notation

A posterior distribution will have a value at each point in its parameter space equal to its probability density at that point. This is not a straightforward value to interpret, so it is common to define credible regions to allow for easier interpretation of results. An $X\%$ credible region contains $X\%$ of the total probability mass of the posterior distribution. There are infinitely many such regions and there is more than one convention as to how to choose it; the definition used in this thesis is that the $X\%$ credible region contains the highest-probability-density $X\%$ of the total probability mass.

A further shorthand used in this thesis is sigma notation, where credible regions are referred to using the analogue of the univariate Gaussian distribution (see Section 2.4.2 below), which contains a given amount of probability mass within N standard deviations of the mean. The conversion between a sigma value of $N\sigma$ and the corresponding credible region is given by

$$N\sigma = \text{erf}\left(\frac{N}{\sqrt{2}}\right) \text{ credible region}, \quad (2.59)$$

where the error function erf is defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}. \quad (2.60)$$

The values for $N = (1, 2, 3, 4, 5)$ are

$$1\sigma = 68.268\,949 \% \text{ credible region}; \quad (2.61)$$

$$2\sigma = 95.449\,974 \% \text{ credible region}; \quad (2.62)$$

$$3\sigma = 99.730\,020 \% \text{ credible region}; \quad (2.63)$$

$$4\sigma = 99.993\,666 \% \text{ credible region}; \quad (2.64)$$

$$5\sigma = 99.999\,943 \% \text{ credible region}, \quad (2.65)$$

but N may take any non-negative value including non-integer values.

2.4.2. Gaussian likelihood and covariance

A common choice of likelihood function is the multivariate Gaussian (also known as normal) distribution. The Gaussian probability density function (PDF) for a data vector x of length k is

$$f_N(x|\mu, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right], \quad (2.66)$$

where $|\cdot|$ denotes a matrix determinant. In Equation (2.66), μ is the mean (expected value) of x ,

$$\langle x \rangle = \mu, \quad (2.67)$$

and Σ is the covariance matrix of x ,

$$\Sigma_{ij} = \text{Cov}(x_i, x_j), \quad (2.68)$$

where covariance is defined as

$$\text{Cov}(x_i, x_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle. \quad (2.69)$$

The suitability of Equation (2.66) as a choice of likelihood function for pseudo- C_ℓ estimates is investigated in Chapter 4, while the covariance of pseudo- C_ℓ estimates is studied in Chapter 5.

Chapter 3

Exact likelihood of pseudo- C_ℓ estimates from Gaussian fields

3.1. Introduction

As described in Chapter 1, the unprecedented precision offered by upcoming weak lensing surveys such as *Euclid* demands that all possible sources of bias in the process of cosmological inference must be examined and controlled. One such source of bias is the likelihood function, which—as described in Chapter 2—is an essential ingredient of Bayesian inference. The traditional, convenient choice throughout cosmology is a multivariate Gaussian likelihood. Depending on the cosmological observable in question, this assumption may be anything from (near-)exact to severely wrong. In the latter case, this error can potentially propagate through to biased cosmological constraints.

This chapter considers the likelihood of observed cosmological power spectra. Even in contemporary analyses these are routinely assumed to be Gaussian distributed (e.g. Hikage et al. 2019; Liu & Madhavacheril 2019; Planck Collaboration et al. 2019), and the choice of covariance matrix in a Gaussian likelihood is an extremely active area of research (e.g. Kodwani, Alonso & Ferreira 2019; Harnois-Déraps, Giblin & Joachimi 2019; Chapter 5). However, it is well known that the true power spectrum likelihood is strongly non-Gaussian, especially on large scales, for weak lensing (Hartlap et al. 2009; Sellentin & Heavens 2018; Sellentin, Heymans & Harnois-Déraps 2018; Hall & Taylor 2022) and also for CMB observations (Percival & Brown 2006; Sun, Wang & Zhan 2013).

For full-sky observations of Gaussian fields with statistically isotropic noise and in the absence of systematic effects, the exact likelihood of observed power spectra is known (Percival & Brown 2006). However, the situation is more complex for realistic observations, and specifically in the presence of a mask. Wandelt, Hivon & Górski (2001) presented the exact distribution of power in a single multipole of a single spin-0 field in the presence of an azimuthally symmetric mask. In this chapter, this is extended to obtain the exact likelihood of an arbitrary number of multipoles of an arbitrary number of correlated spin-0 and spin-2 Gaussian fields, each observed with a mask of any geometry. This is the first mathematically exact approach to such a generalisation. Previously, many approximate forms have been suggested, some of which attempt to model other realistic effects beyond a cut sky (Percival & Brown 2006; Hamimeche & Lewis 2008; Mangilli, Plaszczynski & Tristram 2015; Kalus, Percival & Samushia 2016). Alternative approaches include Gaussianisation of

the data vector (e.g. Wang et al. 2019) or likelihood-free inference (Alsing & Wandelt 2019; Taylor et al. 2019).

This chapter presents the exact distribution of power spectrum estimates from Gaussian fields measured on a cut sky using the pseudo- C_ℓ estimator, which was introduced in Chapter 2. In Section 3.2 the distribution of observed spherical harmonic coefficients—the pseudo- $a_{\ell m}$ s—is derived, demonstrating that their Gaussianity is preserved in the presence of a mask—regardless of its geometry—and exact expressions for the covariance of any pair of pseudo- $a_{\ell m}$ s are presented. In Section 3.3 it is shown that each pseudo- C_ℓ estimator may be written as a quadratic form in the pseudo- $a_{\ell m}$ s, which enables the use of the known joint distribution of such quadratic forms to derive the full, exact joint likelihood of pseudo- C_ℓ estimates from Gaussian fields. It is also shown that the same formalism can be applied to obtain the exact joint likelihood of quadratic maximum likelihood (QML) estimates from Gaussian fields. The QML estimator is an alternative to the pseudo- C_ℓ estimator, which is designed to return minimum-variance estimates provided a suitable choice of fiducial model is made (Tegmark 1997). As an example, Sections 3.4 and 3.5 consider the application to observations of CMB polarisation, though as noted there, these results are equally applicable to large-scale weak lensing data. Section 3.5 demonstrates that the pseudo- C_ℓ likelihood reproduces the exact distribution of EE , BB and EB pseudo- C_ℓ power spectra from Gaussian fields in the presence of a general mask. A discussion of the computational tractability and practical applications is presented along with a summary in Section 3.6.

3.2. Pseudo- $a_{\ell m}$ distribution

In this section the distribution of spherical harmonic coefficients on a general cut sky, the pseudo- $a_{\ell m}$ s, is derived.

3.2.1. Full-sky $a_{\ell m}$ distribution

We begin by considering the distribution of $a_{\ell m}$ s on the full sky. Let us assume correlated, isotropic Gaussian spin-0 or spin-2 cosmological fields $\alpha(\Omega)$, where Ω represents sky coordinates. Greek characters will be used to represent cosmological fields throughout. Each spin-0 field can be decomposed in terms of spherical harmonics as

$$\alpha(\Omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}^{(\alpha)} Y_{\ell m}(\Omega), \quad (3.1)$$

where $a_{\ell m}^{(\alpha)}$ are the spherical harmonic coefficients for field α . Each (complex) spin-2 field can be decomposed in terms of spin-weighted spherical harmonics as

$$\alpha(\Omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(a_{\ell m}^{E(\alpha)} + i a_{\ell m}^{B(\alpha)} \right) {}_2Y_{\ell m}(\Omega), \quad (3.2)$$

$$\alpha^*(\Omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(a_{\ell m}^{E(\alpha)} - i a_{\ell m}^{B(\alpha)} \right) {}_{-2}Y_{\ell m}(\Omega), \quad (3.3)$$

where the superscript * denotes complex conjugation, and $a_{\ell m}^{E(\alpha)}$ and $a_{\ell m}^{B(\alpha)}$ are the spherical harmonic coefficients of the E - and B -mode components of field α , respectively.

For spin-0 and spin-2 spherical harmonics, the $a_{\ell m}$ s have the property that

$$a_{\ell-m}^{(\alpha)} = (-1)^m \left(a_{\ell m}^{(\alpha)} \right)^*, \quad (3.4)$$

which further implies that the $m = 0$ $a_{\ell m}$ s must be real. For a given multipole ℓ , the real and imaginary parts of $a_{\ell m}^{(\alpha)}$ for all $m > 0$ are independently and identically distributed as multivariate Gaussian with mean $\mathbf{0}$ and covariance matrix $\frac{1}{2}\mathbf{C}_\ell$, where the elements of \mathbf{C}_ℓ are given by

$$\mathbf{C}_\ell^{\alpha\beta} = C_\ell^{\alpha\beta}. \quad (3.5)$$

Here, each of α and β can represent either a spin-0 field or the E - or B -mode component of a spin-2 field, and $C_\ell^{\alpha\beta}$ is the underlying angular cross-power spectrum between fields α and β (which may be 0). \mathbf{C}_ℓ is symmetric such that $C_\ell^{\alpha\beta} = C_\ell^{\beta\alpha}$, and $\alpha = \beta$ gives the auto-power spectrum $C_\ell^{\alpha\alpha}$. In the case of $m = 0$, the lack of an imaginary part means that all of the underlying power is in the real part, and hence the $a_{\ell 0}^{(\alpha)}$ are multivariate Gaussian distributed with mean $\mathbf{0}$ and covariance matrix \mathbf{C}_ℓ . The real part of each $a_{\ell m}^{(\alpha)}$ is independent of all imaginary parts, and vice versa. All $a_{\ell m}^{(\alpha)}$ are independent between different ℓ and m ; the only cross-correlation is between $a_{\ell m}^{(\alpha)}$ s having the same ℓ and the same m , but different fields α . The $m < 0$ $a_{\ell m}$ s can be regarded as deterministic functions of their positive- m counterparts following Equation (3.4), rather than separate random variables.

3.2.2. Effect of a cut sky

In real space, the effect of a mask is to multiply each field by a window function $W_\alpha(\Omega)$, which can in general be unique to each field:

$$\tilde{\alpha}(\Omega) = W_\alpha(\Omega) \alpha(\Omega). \quad (3.6)$$

The multiplication in real space is equivalent to a convolution in spherical harmonic space, which has the effect of mixing the $a_{\ell m}$ s between different ℓ , m , and between E - and B -modes in the case of spin-2 fields, to give the pseudo- $a_{\ell m}$ s (Lewis, Challinor & Turok 2001; Brown, Castro & Taylor 2005):

$$\tilde{a}_{\ell m}^{(\alpha)} = \sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{\ell'} {}_0 W_{\ell\ell'}^{mm'} a_{\ell'm'}^{(\alpha)}; \quad (3.7)$$

$$\tilde{a}_{\ell m}^{E(\alpha)} = \sum_{\ell=0}^{\infty} \sum_{m'=-\ell}^{\ell} \left(W_{\ell\ell'mm'}^+ a_{\ell'm'}^{E(\alpha)} + W_{\ell\ell'mm'}^- a_{\ell'm'}^{B(\alpha)} \right); \quad (3.8)$$

$$\tilde{a}_{\ell m}^{B(\alpha)} = \sum_{\ell=0}^{\infty} \sum_{m'=-\ell}^{\ell} \left(W_{\ell\ell'mm'}^+ a_{\ell'm'}^{B(\alpha)} - W_{\ell\ell'mm'}^- a_{\ell'm'}^{E(\alpha)} \right). \quad (3.9)$$

The spin-weighted spherical harmonic space window functions are given by

$${}_s W_{\ell\ell'}^{mm'} = \int d\Omega_s Y_{\ell'm'}(\Omega) W_\alpha(\Omega) {}_s Y_{\ell m}^*(\Omega), \quad (3.10)$$

noting that the optional field-dependence of the mask reflected in $W_\alpha(\Omega)$ also means that each ${}_s W_{\ell\ell'}^{mm'}$ is implicitly field-specific, but the α is dropped on the left-hand side of Equation (3.10) to limit the number of indices. Following Brown, Castro & Taylor (2005), W^+ and W^- are defined as

$$W_{\ell\ell'mm'}^+ = \frac{1}{2} \left({}_2 W_{\ell\ell'}^{mm'} + {}_{-2} W_{\ell\ell'}^{mm'} \right); \quad (3.11)$$

$$W_{\ell\ell'mm'}^- = \frac{i}{2} \left({}_2 W_{\ell\ell'}^{mm'} - {}_{-2} W_{\ell\ell'}^{mm'} \right), \quad (3.12)$$

which may also be specific to each field.

3.2.3. Pseudo- $a_{\ell m}$ distribution

Equations (3.7)–(3.9) describe the effects of a mask on spin-0 and spin-2 fields. More generally, any observed pseudo- $a_{\ell m}$ can be written as a linear combination of full-sky $a_{\ell m}$ s as

$$\tilde{a}_{\ell m}^{(\alpha)} = \sum_{\ell'm'} \sum_{\beta} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'm'}^{(\beta)}} a_{\ell'm'}^{(\beta)}, \quad (3.13)$$

where α and β may each be either a single spin-0 field or the E - or B -mode component of a spin-2 field. This can be expanded into real and imaginary parts as

$$\tilde{a}_{\ell m}^{(\alpha)} = \text{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right) + i \text{Im} \left(\tilde{a}_{\ell m}^{(\alpha)} \right), \quad (3.14)$$

and it is shown in Appendix A.1 that the respective parts are given by

$$\begin{aligned} \text{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right) &= \sum_{\beta, \ell'} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} \right) \text{Re} \left(a_{\ell' 0}^{(\beta)} \right) \right. \\ &\quad + \sum_{m' > 0} \left(\left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right) + (-1)^{m'} \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right) \right] \text{Re} \left(a_{\ell' m'}^{(\beta)} \right) \right. \\ &\quad \left. \left. - \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right) - (-1)^{m'} \text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right) \right] \text{Im} \left(a_{\ell' m'}^{(\beta)} \right) \right); \end{aligned} \quad (3.15)$$

$$\begin{aligned} \text{Im} \left(\tilde{a}_{\ell m}^{(\alpha)} \right) &= \sum_{\beta, \ell'} \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} \right) \text{Re} \left(a_{\ell' 0}^{(\beta)} \right) \right. \\ &\quad + \sum_{m' > 0} \left(\left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right) + (-1)^{m'} \text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right) \right] \text{Re} \left(a_{\ell' m'}^{(\beta)} \right) \right. \\ &\quad \left. \left. + \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right) - (-1)^{m'} \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right) \right] \text{Im} \left(a_{\ell' m'}^{(\beta)} \right) \right); \end{aligned} \quad (3.16)$$

Each derivative in these equations is a constant weight that is uniquely specified by the geometry of the mask, and is independent of each $a_{\ell' m'}^{(\beta)}$. Therefore, Equations (3.15) and (3.16) are simply linear combinations of Gaussian random variables. Any linear combination of Gaussian variables is itself Gaussian distributed, with mean and covariance adding linearly such that

$$Y = \sum_i a_i X_i, \quad X \sim \mathcal{N}(\mu, \Sigma) \quad \Rightarrow \quad Y \sim \mathcal{N} \left(\sum_i a_i \mu_i, \sum_{ii'} a_i a_{i'} \Sigma_{ii'} \right), \quad (3.17)$$

where \sim means ‘distributed as’ and $\mathcal{N}(\mu, \Sigma)$ represents a multivariate Gaussian distribution with mean vector μ and covariance matrix Σ , or for a single variable such as Y , $\mathcal{N}(\mu, \sigma^2)$ represents a univariate Gaussian distribution with mean μ and variance σ^2 . The elements of Σ are given by

$$\Sigma_{ij} = \text{Cov}(X_i, X_j), \quad (3.18)$$

where $\text{Cov}(\cdot, \cdot)$ represents the covariance of two random variables, and the covariance of any variable with itself is its variance. From Equation (3.17) we find that the real and imaginary parts of $\tilde{a}_{\ell m}^{(\alpha)}$ each follow a Gaussian distribution with zero mean, since the full-sky $a_{\ell m}$ s each have zero mean. It is shown in Appendix A.1 that the variance of the real and imaginary part of each pseudo- $a_{\ell m}$ is

$$\begin{aligned} \text{Var} \left(\text{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right) \right) &= \sum_{\beta, \gamma} \sum_{\ell'} C_{\ell'}^{\beta \gamma} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} \right) \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\gamma)}} \right) \right. \\ &\quad + \frac{1}{2} \sum_{m' > 0} \left(\text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right]^* \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\gamma)}} \right) + \text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right]^* \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\gamma)}} \right) \right. \\ &\quad \left. \left. + (-1)^{m'} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\gamma)}} \right) + \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\gamma)}} \right) \right] \right) \right]; \end{aligned} \quad (3.19)$$

$$\begin{aligned} \text{Var} \left(\text{Im} \left(\tilde{a}_{\ell m}^{(\alpha)} \right) \right) &= \sum_{\beta, \gamma} \sum_{\ell'} C_{\ell'}^{\beta \gamma} \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} \right) \text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\gamma)}} \right) \right. \\ &\quad + \frac{1}{2} \sum_{m' > 0} \left(\text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right]^* \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\gamma)}} \right) + \text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right]^* \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\gamma)}} \right) \right. \\ &\quad \left. \left. - (-1)^{m'} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\gamma)}} \right) + \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\gamma)}} \right) \right] \right) \right], \end{aligned} \quad (3.20)$$

where β and γ are all fields correlated with field α , each of which may be either spin-0 or the E - or B -mode component of a spin-2 field.

Since both the real and imaginary parts of each pseudo- $a_{\ell m}$ are Gaussian distributed, the joint distribution for both the real and imaginary parts of all pseudo- $a_{\ell m}$ s for all fields, contained in the vector \tilde{a} , can be described by a multivariate Gaussian distribution,

$$\tilde{a} \sim \mathcal{N} (\mathbf{0}, \Sigma), \quad (3.21)$$

with covariance matrix Σ whose diagonal elements are given by Equations (3.19) and (3.20). The off-diagonal elements can be calculated using the rule—which is not specific to the Gaussian distribution—that the covariance of two linear combinations of random variables is given by

$$\text{Cov} \left(\sum_i \alpha_i X_i, \sum_j \beta_j Y_j \right) = \sum_{ij} \alpha_i \beta_j \text{Cov} (X_i, Y_j), \quad (3.22)$$

where α_i and β_j are constant weights. It is shown in Appendix A.1 that this leads to the following expressions for the elements of Σ :

$$\begin{aligned} \text{Cov} \left(\text{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right), \text{Re} \left(\tilde{a}_{\ell' m'}^{(\beta)} \right) \right) = & \sum_{\gamma, \varepsilon} \sum_{\ell''} C_{\ell''}^{\gamma \varepsilon} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' 0}^{(\gamma)}} \right) \text{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' 0}^{(\varepsilon)}} \right) \right. \\ & + \frac{1}{2} \sum_{m'' > 0} \left(\text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right]^* \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) + \text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right]^* \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) \right. \\ & \left. \left. + (-1)^{m''} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) + \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) \right] \right); \end{aligned} \quad (3.23)$$

$$\begin{aligned} \text{Cov} \left(\text{Im} \left(\tilde{a}_{\ell m}^{(\alpha)} \right), \text{Im} \left(\tilde{a}_{\ell' m'}^{(\beta)} \right) \right) = & \sum_{\gamma, \varepsilon} \sum_{\ell''} C_{\ell''}^{\gamma \varepsilon} \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' 0}^{(\gamma)}} \right) \text{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' 0}^{(\varepsilon)}} \right) \right. \\ & + \frac{1}{2} \sum_{m'' > 0} \left(\text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right]^* \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) + \text{Re} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right]^* \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) \right. \\ & \left. \left. - (-1)^{m''} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) + \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) \right] \right); \end{aligned} \quad (3.24)$$

$$\begin{aligned} \text{Cov} \left(\text{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right), \text{Im} \left(\tilde{a}_{\ell' m'}^{(\beta)} \right) \right) = & \sum_{\gamma, \varepsilon} \sum_{\ell''} C_{\ell''}^{\gamma \varepsilon} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' 0}^{(\gamma)}} \right) \text{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' 0}^{(\varepsilon)}} \right) \right. \\ & + \frac{1}{2} \sum_{m'' > 0} \left(\text{Im} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right]^* \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) + \text{Im} \left(\left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right]^* \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) \right. \\ & \left. \left. + (-1)^{m''} \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) + \text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) \right] \right). \end{aligned} \quad (3.25)$$

This is the first key result of this chapter: that the spherical harmonic coefficients of correlated Gaussian fields measured on a cut sky, which follow a multivariate Gaussian distribution, have mean $\boldsymbol{0}$ and covariance matrix Σ whose elements are given by Equations (3.23)–(3.25). For a given cosmological model and survey mask, these expressions can then be applied to a particular set of observational probes by specifying the fields in question (α, β), the cosmological signals ($C_{\ell}^{\gamma \varepsilon}$), and all non-zero derivative terms, which once again are completely specified by the geometry of the survey mask. This is demonstrated explicitly for the example of the CMB later, in Section 3.4.1.

3.3. Pseudo- C_ℓ distribution

In this section, the distribution of pseudo- C_ℓ estimates for auto- and cross-power spectra of an arbitrary number of correlated spin-0 and spin-2 Gaussian fields is derived for a general

cut sky.

3.3.1. The pseudo- C_ℓ estimator

On the full sky, the covariance between either real parts or imaginary parts of two $a_{\ell m}$ s having the same (ℓ, m) is equal to the corresponding underlying power spectrum, modulo a factor of $\frac{1}{2}$ for $m > 0$ (see Section 3.2.1). Therefore, an unbiased estimator of the power spectrum is given by an appropriately weighted sample covariance of observed $a_{\ell m}$ s:

$$\begin{aligned}\widehat{C}_\ell^{\alpha\beta} &= \frac{1}{2\ell+1} \left[\operatorname{Re}(a_{\ell 0}^{(\alpha)}) \operatorname{Re}(a_{\ell 0}^{(\beta)}) + 2 \sum_{m=1}^{\ell} [\operatorname{Re}(a_{\ell m}^{(\alpha)}) \operatorname{Re}(a_{\ell m}^{(\beta)}) + \operatorname{Im}(a_{\ell m}^{(\alpha)}) \operatorname{Im}(a_{\ell m}^{(\beta)})] \right] \\ &= \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m}^{(\alpha)} (a_{\ell m}^{(\beta)})^*,\end{aligned}\tag{3.26}$$

where we implicitly take the real part of the result. The auto-spectrum estimator is given by the special case where $\alpha = \beta$. Under a cut sky, this becomes the pseudo- C_ℓ estimator (Wandelt, Hivon & Górski 2001; Hivon et al. 2002; Brown, Castro & Taylor 2005), written in terms of the pseudo- $a_{\ell m}$ s:

$$\widetilde{C}_\ell^{\alpha\beta} = \frac{1}{2\ell+1} \left[\operatorname{Re}(\tilde{a}_{\ell 0}^{(\alpha)}) \operatorname{Re}(\tilde{a}_{\ell 0}^{(\beta)}) + 2 \sum_{m=1}^{\ell} [\operatorname{Re}(\tilde{a}_{\ell m}^{(\alpha)}) \operatorname{Re}(\tilde{a}_{\ell m}^{(\beta)}) + \operatorname{Im}(\tilde{a}_{\ell m}^{(\alpha)}) \operatorname{Im}(\tilde{a}_{\ell m}^{(\beta)})] \right].\tag{3.27}$$

This can be written as a matrix equation involving the vector of all pseudo- $a_{\ell m}$ s, $\tilde{\mathbf{a}}$:

$$\widetilde{C}_\ell^{\alpha\beta} = \tilde{\mathbf{a}}^\top \mathbf{M}_\ell^{\alpha\beta} \tilde{\mathbf{a}},\tag{3.28}$$

where $\mathbf{M}_\ell^{\alpha\beta}$ is a real symmetric matrix chosen to pick out the correct elements of $\tilde{\mathbf{a}}$ to match the expression in Equation (3.27). Since this may not be obvious, it is here demonstrated by choosing an order for $\tilde{\mathbf{a}}$ and seeing the corresponding form for $\mathbf{M}_\ell^{\alpha\beta}$. However, it should be noted that this order is arbitrary and could be chosen, for example, to optimise computation of the likelihood. For N correlated fields, each of which is either a spin-0 field or the E - or B -mode component of a spin-2 field, and a maximum measured multipole of ℓ_{\max} , $\tilde{\mathbf{a}}$ can be decomposed into three hierarchical levels as

$$\tilde{\mathbf{a}} = \begin{pmatrix} \tilde{\mathbf{a}}^{(\text{field 1})} \\ \tilde{\mathbf{a}}^{(\text{field 2})} \\ \vdots \\ \tilde{\mathbf{a}}^{(\text{field } N)} \end{pmatrix}; \quad \tilde{\mathbf{a}}^{(\text{field } \alpha)} = \begin{pmatrix} \tilde{\mathbf{a}}_{\ell=0}^{(\alpha)} \\ \tilde{\mathbf{a}}_{\ell=1}^{(\alpha)} \\ \vdots \\ \tilde{\mathbf{a}}_{\ell_{\max}}^{(\alpha)} \end{pmatrix}; \quad \tilde{\mathbf{a}}_\ell^{(\alpha)} = \begin{pmatrix} \text{Re} \left(\tilde{a}_{\ell 0}^{(\alpha)} \right) \\ \text{Re} \left(\tilde{a}_{\ell 1}^{(\alpha)} \right) \\ \text{Im} \left(\tilde{a}_{\ell 1}^{(\alpha)} \right) \\ \text{Re} \left(\tilde{a}_{\ell 2}^{(\alpha)} \right) \\ \text{Im} \left(\tilde{a}_{\ell 2}^{(\alpha)} \right) \\ \vdots \\ \text{Re} \left(\tilde{a}_{\ell \ell}^{(\alpha)} \right) \\ \text{Im} \left(\tilde{a}_{\ell \ell}^{(\alpha)} \right) \end{pmatrix}. \quad (3.29)$$

We can then similarly decompose $\mathbf{M}_\ell^{\alpha\beta}$ into blocks to pick out the correct elements:

$$\mathbf{M}_\ell^{\alpha\beta} = \begin{array}{c} \text{Field :} \\ \begin{matrix} & 1 & 2 & \dots & \alpha & \dots & \beta & \dots & N \\ 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ 2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{M}_\ell & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta & \mathbf{0} & \mathbf{0} & \dots & \mathbf{M}_\ell & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ N & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \end{matrix} \end{array}; \quad (3.30)$$

$$\mathbf{M}_\ell = \begin{array}{c} \ell' : \\ \begin{matrix} & 0 & 1 & \dots & \ell & \dots & \ell_{\max} \\ 0 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \ell & \mathbf{0} & \mathbf{0} & \dots & \mathbf{M} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \ell_{\max} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \end{matrix} \end{array}; \quad (3.31)$$

$$\mathbf{M} = \left(\frac{1 + \delta_{\alpha\beta}}{2\ell + 1} \right) \text{diag} \left(\frac{1}{2}, 1, 1, 1, \dots \right). \quad (3.32)$$

The Kronecker delta $\delta_{\alpha\beta}$ accounts for the auto-spectrum case where $\alpha = \beta$. \mathbf{M} has $2\ell + 1$ diagonal elements, so the size of \mathbf{M} is dependent on ℓ but its form is not. Each block in \mathbf{M}_ℓ has $2\ell' + 1$ elements in each dimension, so \mathbf{M}_ℓ has a total of $\sum_{\ell'=0}^{\ell_{\max}} (2\ell' + 1) = (\ell_{\max} + 1)^2$ elements along each side. This is also true of every zero block in $\mathbf{M}_\ell^{\alpha\beta}$, so $\mathbf{M}_\ell^{\alpha\beta}$ has a total of $N \times (\ell_{\max} + 1)^2$ elements along each side. The choice of $\ell_{\min} = 0$ is arbitrary, and any other ℓ_{\min} could be chosen as long as this choice is consistent between $\mathbf{M}_\ell^{\alpha\beta}$ and $\tilde{\mathbf{a}}$.

3.3.2. The joint distribution of quadratic forms

Each cut-sky cross- or auto-spectrum estimator can be written in the form of Equation (3.28), each with the same multivariate Gaussian vector $\tilde{\alpha}$ and different symmetric matrix $\mathbf{M}_\ell^{\alpha\beta}$ depending on the field(s) and multipole in question. The joint characteristic function (CF) of an arbitrary number of such quadratic forms, $\varphi(\mathbf{t})$, has a known analytic form (Good 1963):

$$\varphi(\mathbf{t}) = \left| \mathbf{I} - 2i \sum_{\ell} \sum_{\alpha\beta} t_\ell^{\alpha\beta} \mathbf{M}_\ell^{\alpha\beta} \Sigma \right|^{-1/2}, \quad (3.33)$$

where $|\cdot|$ denotes the matrix determinant. The joint CF is defined as the expectation value $\langle e^{i\mathbf{t}\cdot\tilde{\mathbf{C}}}\rangle$, which can also be written as a Fourier integral, i.e.

$$\varphi(\mathbf{t}) = \int d\tilde{\mathbf{C}} \exp(i\mathbf{t}\cdot\tilde{\mathbf{C}}) f(\tilde{\mathbf{C}}), \quad (3.34)$$

where each $\{\tilde{C}_\ell^{\alpha\beta}, t_\ell^{\alpha\beta}\}$ are a Fourier pair¹ and $\tilde{\mathbf{C}}$ and \mathbf{t} are the respective vectors of all $\tilde{C}_\ell^{\alpha\beta}$ and $t_\ell^{\alpha\beta}$ for all fields and multipoles in question. $f(\tilde{\mathbf{C}})$ is the joint probability density function of all observed pseudo- C_ℓ s. Equation (3.34) may then be inverted to yield the exact joint distribution of pseudo- C_ℓ estimates from Gaussian fields:

$$f(\tilde{\mathbf{C}}) = \frac{1}{(2\pi)^v} \int d\mathbf{t} \exp(-i\tilde{\mathbf{C}}\cdot\mathbf{t}) \left| \mathbf{I} - 2i \sum_{\ell} \sum_{\alpha\beta} t_\ell^{\alpha\beta} \mathbf{M}_\ell^{\alpha\beta} \Sigma \right|^{-1/2}. \quad (3.35)$$

Here v is the length of the data vector, which is equal to $N \times (\ell_{\max} + 1)^2$ when considering N fields and all multipoles from 0 to ℓ_{\max} . Equation (3.35) is normalised by construction, such that it integrates to 1 over the range of possible $\tilde{\mathbf{C}}$.

Equation (3.35) is the second key result of this chapter: the exact joint likelihood of an arbitrary number of auto- and cross-pseudo- C_ℓ estimates from an arbitrary number of correlated spin-0 and spin-2 Gaussian fields. It implicitly depends on the cosmological model via the covariance matrix of pseudo- $a_{\ell m}$ s, Σ , whose elements are given in Equations (3.23)–(3.25). These elements depend on both the underlying power spectra and the mask for each field.

Equation (3.35) is related to Equation (16) of Wandelt, Hivon & Górski (2001) and to Equation (C6) of Hamimeche & Lewis (2008), as all three are instances of the general distribution of quadratic forms. However, Equation (3.35) presented here is a much more generally applicable result than Equation (16) of Wandelt, Hivon & Górski (2001), which is restricted to a single multipole of a single spin-0 power spectrum observed with an azimuthally symmetric mask. Equation (C6) of Hamimeche & Lewis (2008), meanwhile, describes the

¹ This should not be confused with the Fourier pair of $\{\ell, \theta\}$ which relate the power spectrum C_ℓ to the correlation function $\xi(\theta)$.

distribution of a single multipole of a full-sky weighted cross-spectrum estimator.

3.3.2.1. The likelihood of a subset of pseudo- C_ℓ estimates

The formalism presented in Equations (3.29)–(3.32) allows one to obtain the full multivariate likelihood for all correlated fields and multipoles from ℓ_{\min} to ℓ_{\max} . However, this likelihood may be adapted depending on exactly which pseudo- C_ℓ s are required, and only the elements of Σ that contribute to the relevant estimators must be calculated. For example, to calculate the joint likelihood of $\tilde{C}_{\ell=2}^{\alpha\alpha}$ and $\tilde{C}_{\ell=4}^{\alpha\alpha}$, the pseudo- $a_{\ell m}$ vector would be

$$\tilde{\mathbf{a}} = (\operatorname{Re} \tilde{a}_{20}, \operatorname{Re} \tilde{a}_{21}, \operatorname{Im} \tilde{a}_{21}, \operatorname{Re} \tilde{a}_{22}, \operatorname{Im} \tilde{a}_{22}, \operatorname{Re} \tilde{a}_{40}, \operatorname{Re} \tilde{a}_{41}, \\ \operatorname{Im} \tilde{a}_{41}, \operatorname{Re} \tilde{a}_{42}, \operatorname{Im} \tilde{a}_{42}, \operatorname{Re} \tilde{a}_{43}, \operatorname{Im} \tilde{a}_{43}, \operatorname{Re} \tilde{a}_{44}, \operatorname{Im} \tilde{a}_{44})^\top, \quad (3.36)$$

where the (α) subscripts have been omitted for clarity. $\mathbf{M}_2^{\alpha\alpha}$ and $\mathbf{M}_4^{\alpha\alpha}$ could then be chosen to pick out the appropriate elements of $\tilde{\mathbf{a}}$.

It is mathematically straightforward to extend this formalism to obtain the exact distribution of deconvolved power estimates from Gaussian fields, as this is a linear operation in the pseudo- C_ℓ s and hence a quadratic form in the pseudo- $a_{\ell m}$ s. However, as discussed in Chapter 2, this approach offers no additional constraining power and introduces additional possible sources of error, so is not discussed further in this chapter. Similarly, the extension to obtain the distribution of bandpowers through linear binning of multipoles²—which is often necessary if one wishes to obtain deconvolved power estimates—is straightforward, requiring only appropriate changes to the selection matrices $\mathbf{M}_\ell^{\alpha\beta}$.

3.3.2.2. The likelihood of Quadratic Maximum Likelihood estimates

At low ℓ one may prefer to use a Quadratic Maximum Likelihood (QML) power spectrum estimator (Tegmark 1997), due to its optimality. The price is an increased computational cost, and diminishing returns at higher multipoles, compared to the pseudo- C_ℓ estimator (Efstathiou 2004). By design, the QML estimator is also a quadratic form:

$$y_\ell^{\alpha\beta} = \mathbf{x}^\top \mathbf{E}_\ell^{\alpha\beta} \mathbf{x}, \quad (3.37)$$

with

$$\mathbf{E}_\ell^{\alpha\beta} = \frac{1}{2} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial C_\ell^{\alpha\beta}} \mathbf{C}^{-1}, \quad (3.38)$$

where \mathbf{x} is the vector of all pixel values, which is multivariate Gaussian distributed with covariance \mathbf{C} . The y_ℓ may be scaled and linearly transformed to provide unbiased estimates of the underlying power spectrum, but—beyond the iterative procedure described below, and as with pseudo- C_ℓ estimates—this step is unnecessary for cosmological inference and

² ‘Linear’ in the sense that each bandpower is a linear combination of C_ℓ s, not that the distribution of bandpowers across angular scales is linear.

only the distribution of y_ℓ values is discussed here.

Since both \mathbf{C} and $C_\ell^{\alpha\beta}$, and hence also $\mathbf{E}_\ell^{\alpha\beta}$, depend on the true power spectrum which is itself being estimated by this process, an iterative procedure is necessary. Typically, one chooses an initial set of power spectra, which are used to calculate the ingredients to $\mathbf{E}_\ell^{\alpha\beta}$ and to thereby generate estimates for the true power spectra of the data via y_ℓ , which are then used to recalculate \mathbf{C} and $C_\ell^{\alpha\beta}$, and the process is repeated until a level of convergence is reached. Multiple starting points may be used to check robustness to the choice of initial power spectra. (See e.g. Bond, Jaffe & Knox 1998, for more details).

From Equation (3.37) and the general distribution of quadratic forms, it is straightforward to write down the joint likelihood of a set of QML estimates \mathbf{y} ,

$$f(\mathbf{y}) = \frac{1}{(2\pi)^\nu} \int dt \exp(-i\mathbf{y} \cdot \mathbf{t}) \left| \mathbf{I} - 2i \sum_{\ell} \sum_{\alpha\beta} t_\ell^{\alpha\beta} \mathbf{E}_\ell^{\alpha\beta} \mathbf{C} \right|^{-1/2}, \quad (3.39)$$

where ν remains the length of the data vector, which is now equal to the total number of pixels across all maps. A potentially useful property of Equation (3.39) is that the matrices needed to evaluate it— \mathbf{C} and the set of $\mathbf{E}_\ell^{\alpha\beta}$ —are the same matrices needed to evaluate the estimator itself, reducing the amount of additional work needed to evaluate the likelihood.

Theoretically, one could also obtain the joint distribution of both pseudo- C_ℓ and QML estimates. This would require writing both as quadratic forms in the same underlying quantity, such as the full-sky $a_{\ell m s}$. This is mathematically possible because the transforms from full-sky $a_{\ell m s}$ to both pixels and cut-sky pseudo- $a_{\ell m s}$ are linear, meaning that a quadratic form in one can also be written as a quadratic form in the other. These transforms could then be encoded in the selection matrix in place of $\mathbf{M}_\ell^{\alpha\beta}$ in Equation (3.35) or $\mathbf{E}_\ell^{\alpha\beta}$ in Equation (3.39), with the comparatively simple full-sky $a_{\ell m}$ covariance matrix in place of Σ or \mathbf{C} respectively. However, in practice this is unlikely to be an attractive or computationally tractable option. An alternative would be to form the approximate joint distribution using a Gaussian copula, which would avoid the need to write both sets of estimates in terms of a common basis. Instead, it would require the correlation coefficient or covariance between QML and pseudo- C_ℓ estimates, expressions for which are given in Efstathiou (2004).

3.4. Application to cosmic microwave background polarisation

In this section it is demonstrated that the pseudo- C_ℓ likelihood presented in Section 3.3 is correct by calculating the full joint distribution of three multipoles of the CMB polarisation power spectra, and comparing to the observed distribution obtained from simulations. It is shown that the likelihood exactly reproduces correlations between spectra. It also naturally models correlations between any number of multipoles, but here three-dimensional

distributions are considered, to limit the computational resources needed to calculate each full distribution.

The likelihood is equally applicable to weak lensing, which is the main focus of this thesis, but the CMB has been chosen here to highlight the full capture by the likelihood of leakage between E - and B -modes due to a cut sky. While this is of interest in weak lensing observations too, it is essential to account for in CMB experiments searching for evidence of inflation in the low- ℓ BB power spectrum, if the small signal is to be uncovered beneath the leakage from the much larger E -mode power. Focusing on low ℓ (here $\ell = 2$) allows the computational requirements to be kept to a minimum, since the full pseudo- $a_{\ell m}$ covariance is required to be calculated, and there are $v = 2\ell + 1$ $a_{\ell m}$ s per multipole.

3.4.1. CMB pseudo- $a_{\ell m}$ covariance

In the case of the CMB, the relevant fields are the temperature field, the E -mode polarisation and the B -mode polarisation. We therefore identify

$$a_{\ell m}^{(\alpha)} \in \{T_{\ell m}, E_{\ell m}, B_{\ell m}\} \quad (3.40)$$

$$\tilde{a}_{\ell m}^{(\alpha)} \in \{\tilde{T}_{\ell m}, \tilde{E}_{\ell m}, \tilde{B}_{\ell m}\}. \quad (3.41)$$

The fields are mixed as (Lewis, Challinor & Turok 2001; Brown, Castro & Taylor 2005)

$$\tilde{T}_{\ell m} = \sum_{\ell' m'} {}_0 W_{\ell \ell'}^{mm'} T_{\ell' m'}; \quad (3.42)$$

$$\tilde{E}_{\ell m} = \sum_{\ell' m'} (W_{\ell \ell' mm'}^+ E_{\ell' m'} + W_{\ell \ell' mm'}^- B_{\ell' m'}); \quad (3.43)$$

$$\tilde{B}_{\ell m} = \sum_{\ell' m'} (W_{\ell \ell' mm'}^+ B_{\ell' m'} - W_{\ell \ell' mm'}^- E_{\ell' m'}), \quad (3.44)$$

giving the relevant derivatives

$$\frac{\partial \tilde{T}_{\ell m}}{\partial T_{\ell' m'}} = {}_0 W_{\ell \ell'}^{mm'}; \quad (3.45)$$

$$\frac{\partial \tilde{E}_{\ell m}}{\partial E_{\ell' m'}} = W_{\ell \ell' mm'}^+; \quad \frac{\partial \tilde{E}_{\ell m}}{\partial B_{\ell' m'}} = W_{\ell \ell' mm'}^-; \quad (3.46)$$

$$\frac{\partial \tilde{B}_{\ell m}}{\partial E_{\ell' m'}} = -W_{\ell \ell' mm'}^-; \quad \frac{\partial \tilde{B}_{\ell m}}{\partial B_{\ell' m'}} = W_{\ell \ell' mm'}^+. \quad (3.47)$$

Here only the polarisation fields are considered, as CMB polarisation is the focus of many current and future experiments searching for evidence of inflation (e.g. Hui et al. 2018; Ade et al. 2019; Abazajian et al. 2016; Hazumi et al. 2019). However, this formalism naturally extends to include the temperature field, as well as including cross-correlation between fields observed by different detectors. Inserting the derivatives in Equations (3.45)–(3.47) into the

general pseudo- $a_{\ell m}$ covariance matrix elements given in Equations (3.23)–(3.25) gives the elements of the CMB polarisation pseudo- $a_{\ell m}$ covariance matrix. Here it is assumed that all underlying C_ℓ^{EB} vanish, as is the case in any model in which parity is conserved, but Equations (3.23)–(3.25) naturally allow for non-zero underlying C_ℓ^{EB} .

3.4.2. Implementation

The calculation of the pseudo- $a_{\ell m}$ covariance matrix requires theory power spectra and harmonic space window functions. Here, CAMB (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012) was used through the CosmoSIS³ interface (Zuntz et al. 2015) to generate CMB power spectra using the Λ CDM+ r model with *Planck* 2018 best fit parameters from Planck Collaboration et al. (2020b) and tensor-to-scalar ratio $r = 0.01$. The harmonic space window function transform in Equation (3.10) is manually implemented using spin-weighted spherical harmonics provided by the `spherical_functions`⁴ Python package (Boyle, Stein & Gross 2021). However, we found that these only work reliably up to $\ell \approx 35$, so for the tests in this section $\ell_{\max} = 30$ is imposed. Alternative implementations could include calculation of the harmonic space window functions by expressing them in terms of Wigner $3j$ symbols (Hivon et al. 2002) and making use of their recursion relations (Lewis, Challinor & Turok 2001; see also the Appendices of Hamimeche & Lewis 2008, 2009).

When calculating the joint characteristic function in Equation (3.33), we found that it is more numerically stable to use an alternative form that avoids the need to evaluate the determinant of a complex matrix:

$$\varphi(\mathbf{t}) = \prod_j (1 - 2i\lambda_j)^{-1/2}, \quad \lambda_j \in \lambda \left(\sum_\ell \sum_{\alpha\beta} t_\ell^{\alpha\beta} \mathbf{M}_\ell^{\alpha\beta} \boldsymbol{\Sigma} \right); \quad (3.48)$$

i.e., the product is over all eigenvalues of the real matrix $\sum_\ell \sum_{\alpha\beta} t_\ell^{\alpha\beta} \mathbf{M}_\ell^{\alpha\beta} \boldsymbol{\Sigma}$. It is shown in Appendix A.2 that this form is mathematically equivalent to the form presented in Equation (3.33). To calculate the full joint likelihood distribution from the joint CF, Equation (3.35) may be written in terms of a Fast Fourier Transform (FFT):

$$\begin{aligned} f(\tilde{\mathbf{C}}) &= \frac{1}{(2\pi)^n} \int_n \left(\prod_k dt_k \right) \exp \left(-i \sum_k \tilde{C}_k t_k \right) \varphi(\mathbf{t}) \\ &= \lim_{\Delta t_k \rightarrow 0} \frac{1}{(2\pi)^n} \left(\prod_k \Delta t_k \right) \exp \left(-i \sum_k t_{k0} \tilde{C}_k \right) \text{FFT}^n [\varphi(\mathbf{t})], \end{aligned} \quad (3.49)$$

where the index k is used as a proxy to encapsulate all summation variables (ℓ, α, β) in Equation (3.48). We used the NumPy⁵ n -dimensional FFT (Harris et al. 2020):

³ <https://bitbucket.org/joezuntz/cosmosis>

⁴ https://github.com/moble/spherical_functions

⁵ <https://numpy.org>

$$\text{FFT}^n[f(t)](x) = \prod_k \sum_{m_k=1}^{N_k} f(\{t_{k0} + m_k \Delta t_k\}) \exp\left(-i \sum_{k'} \frac{x_{k'} m_{k'} \Delta t_{k'}}{N_{k'}}\right), \quad (3.50)$$

where each t_k is discretised as $t_k = t_{k0} + m_k \Delta t_k$.

3.4.3. Simulations

We generated 276 million realisations of the CMB polarisation field from the fixed theory power spectra using the `synfast` routine in `HEALPix`⁶ (Górski et al. 2005; Zonca et al. 2019) with a resolution of `nside` = 128. To match the assumptions made in the theoretical distribution, a value of $\ell_{\max} = 30$ is also imposed in the input power. In general, the accuracy of a finite ℓ_{\max} requires either a band-limited signal (such as is the case for the CMB), a well-behaved mask or apodisation to strongly suppress long-range mode mixing. For each realisation, we applied the polarisation field mask used in the WMAP 9-year analysis, corresponding to a sky fraction of 73.2 %. The mask is described further in Bennett et al. (2013). The EE , BB and EB pseudo- C_ℓ power spectra were measured from the masked maps using `anafast`. We formed marginal histograms of the power at $\ell = 2, 5$ and 10 from each spectrum and a three-dimensional histogram of the measured $\ell = 2$ power from all three spectra, which were selected to demonstrate the ability of the exact likelihood to exactly describe correlations between spectra when applied to Gaussian fields. It also naturally (and exactly, for Gaussian fields) describes correlations between multipoles, though this is not shown here to limit the computational complexity and number of required simulations. The results shown here have been chosen to focus on the fact that the likelihood exactly models the mixing of power between E - and B -modes of Gaussian fields, as this is a particular difficulty for approximate methods due to the large discrepancy in the underlying E - and B -mode power. The ability of the likelihood to exactly describe correlations between multipoles, for Gaussian fields, was separately tested and confirmed.

3.4.4. Comparison to approximation

Let us also consider an approximation to the exact likelihood for Gaussian fields, to demonstrate the importance of accurately modelling the full multidimensional distribution. On the full sky, the joint distribution of \hat{C}_ℓ^{EE} , \hat{C}_ℓ^{BB} and \hat{C}_ℓ^{EB} for a single fixed multipole ℓ follows a Wishart distribution (e.g. Percival & Brown 2006; Chapter 4):

$$\begin{pmatrix} \hat{C}_\ell^{EE} & \hat{C}_\ell^{EB} \\ \hat{C}_\ell^{EB} & \hat{C}_\ell^{BB} \end{pmatrix} \sim W_2(\nu, \mathbf{W}_\ell), \quad \nu = 2\ell + 1, \quad \mathbf{W}_\ell = \frac{1}{2\ell + 1} \begin{pmatrix} C_\ell^{EE} & 0 \\ 0 & C_\ell^{BB} \end{pmatrix}, \quad (3.51)$$

⁶ <https://healpix.sourceforge.io>

where $W_p(d, \mathbf{S})$ denotes the rank- p Wishart distribution with d degrees of freedom and scale matrix \mathbf{S} . We may adapt this for the cut sky using ‘effective’ parameters of ν_{eff} and $\mathbf{W}_\ell^{\text{eff}}$. The values of these effective parameters are chosen to be those that best fit the marginal distributions of \tilde{C}_ℓ^{EE} , \tilde{C}_ℓ^{BB} and \tilde{C}_ℓ^{EB} , as calculated with the exact likelihood, assuming Gaussian fields. This is not intended to be taken as a viable or attractive alternative to the exact likelihood; on the contrary, it is included to demonstrate that while a given approximation may fit the marginals of the likelihood well, it can still fail to capture the details of the full multidimensional distribution. In addition to this, the Wishart approximation can only model correlations between different spectra for the same ℓ , and not correlations between multipoles. The exact likelihood for Gaussian fields presented in this chapter, in contrast, exactly models correlations not only between spectra but also between multipoles, a behaviour which has been verified separately to the tests presented in this chapter. However, as discussed above, this example has been chosen to focus on the former behaviour.

A common alternative is to use a multivariate Gaussian approximation to the likelihood. With a suitable choice of mean, the problem then reduces to finding the most appropriate choice of covariance matrix, which can be calibrated with simulations (see Chapter 5). However, at low multipoles the true likelihood is so skewed that a Gaussian is an extremely poor approximation, which can lead to biased results and, if not modified, can assign non-zero probability to unphysical results such as negative auto-power spectra. For this reason, the 2018 *Planck* analysis uses a pixel-based likelihood below $\ell = 30$ as described in Planck Collaboration et al. (2019). (Later work, presented in Chapter 4, reveals that despite these shortcomings, a Gaussian likelihood is sufficient to obtain accurate results from weak lensing power spectra.)

3.5. Results

3.5.1. Marginal distributions

Figure 3.1 shows the marginal distributions of \tilde{C}_ℓ^{EE} , \tilde{C}_ℓ^{BB} and \tilde{C}_ℓ^{EB} for each of $\ell = 2, 5$ and 10 , to compare the prediction of the exact likelihood for Gaussian fields to the distributions observed in the simulations. Each histogram uses 300 bins. There is an excellent fit between the predicted and observed distribution, with no visible noise in the histograms due to the large number of events in each marginal distribution. The predicted likelihood exactly reproduces both the shape and amplitude of the observed distributions, including the considerable skewness in the auto-spectra. This skewness is reduced for higher multipoles, which is consistent with the full-sky behaviour of the likelihood. Each auto- \tilde{C}_ℓ in Figure 3.1 has been scaled by the relevant theory C_ℓ used to generate both the theoretical likelihood and the simulations. In the case of the cross-spectrum \tilde{C}_ℓ^{EB} , there is no input C_ℓ^{EB}

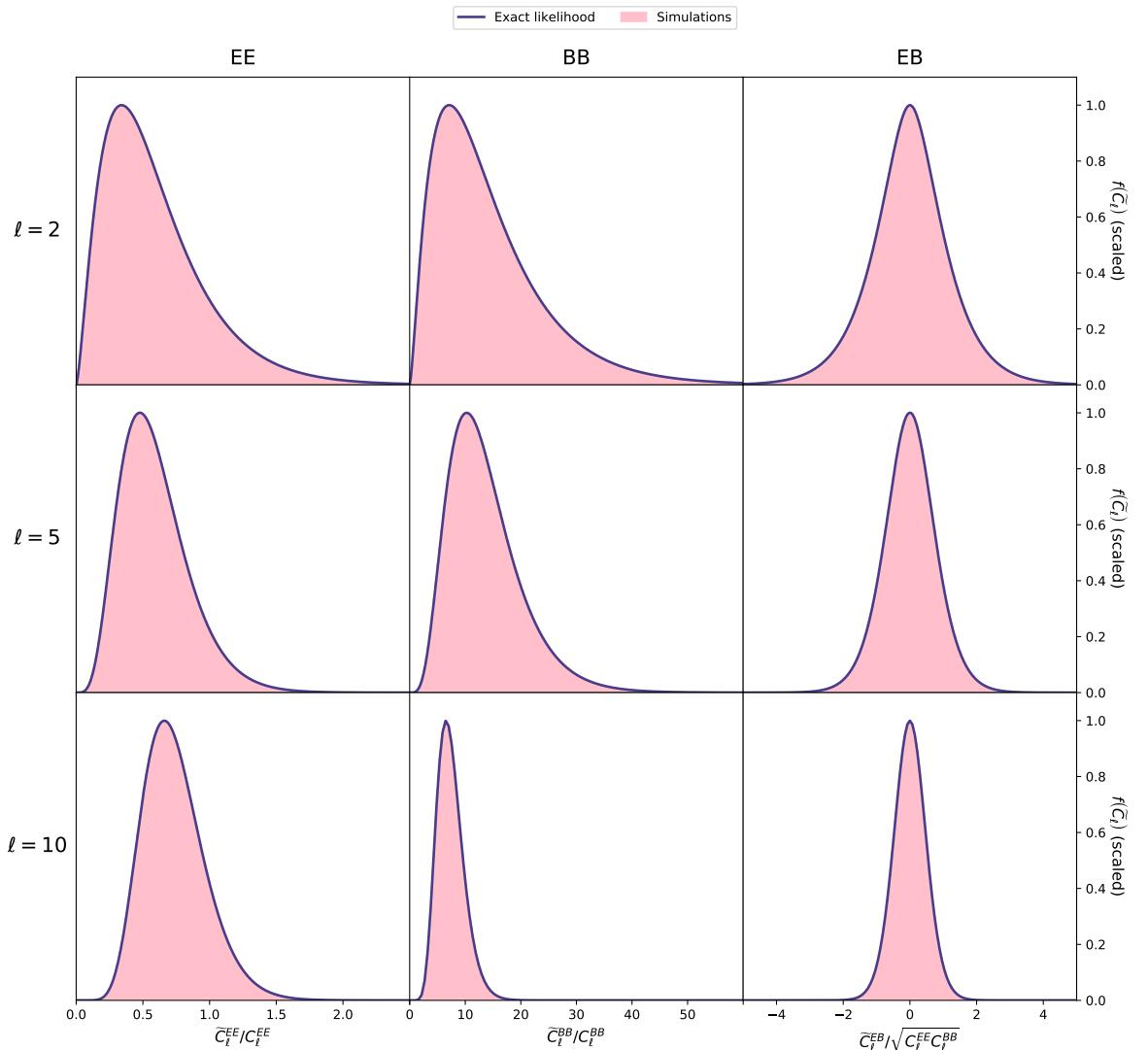


Figure 3.1. The marginal distributions of \tilde{C}_ℓ^{EE} , \tilde{C}_ℓ^{BB} and \tilde{C}_ℓ^{EB} for $\ell = 2, 5$ and 10 , predicted by the exact likelihood for Gaussian fields presented in this chapter (blue curves) compared to those observed in the simulations (pink histograms). The maximum value of each curve has been rescaled to 1 for ease of comparison.

so instead $\sqrt{C_\ell^{EE} C_\ell^{BB}}$ is used for the normalisation. This scaling allows us to observe that the E -mode power is reduced by the sky cut while the B -mode power is increased. This is a result of E - B mixing: the EE power spectrum is much larger in magnitude than the BB power spectrum (by a factor ~ 200 at $\ell = 2$), meaning that the E - B mixing induced by the mask leads to a relative increase in B -mode power at the expense of E -mode power.

3.5.2. Correlation between spectra

As well as exactly reproducing marginal distributions, the exact likelihood naturally describes correlations both between multipoles of the same spectrum and between spectra, for Gaussian fields. As described in Section 3.4, we formed the three-dimensional joint likelihood of \tilde{C}_ℓ^{EE} , \tilde{C}_ℓ^{BB} and \tilde{C}_ℓ^{EB} for $\ell = 2$. We formed the corresponding simulated distribution by binning events in three dimensions, using 100 bins in each dimension. We then integrated the exact likelihood over the volume of each histogram bin to allow for comparison between theory and simulations. Figures 3.2–3.4 show two-dimensional slices of this three-dimensional likelihood. Each slice corresponds to fixing a single histogram bin in one dimension and shows all bins in the other two dimensions. The exact likelihood appears to accurately match the observed distributions in all six slices to within pixel noise that arises from the finite number of realisations in the simulations. The right-hand panel for each slice shows the logarithmic fractional residual, defined as

$$r = \log_{10} \left(\frac{|\text{sampled density from simulations} - \text{density from exact likelihood}|}{\text{density from exact likelihood}} \right). \quad (3.52)$$

Bins with no sampled events have $r = 0$ and appear as white in Figures 3.2–3.4. These areas were not explicitly excluded, but their probability is very low (albeit non-zero). No clear evidence of structure is otherwise seen in these residuals, indicating that these bins contain only noise. This is mostly at the level of $r \approx -4$ to -2 , except for a small number of outlying bins whose probability density is so low that the expected number of events in each bin from the 276 million simulations is significantly less than 1, leading to fractional residual values up to $r \approx 2$ in those bins in which an event was observed.

3.5.2.1. Comparison to approximation

In this section, the exact likelihood is compared to a Wishart distribution with fitted parameters ν_{eff} and $\mathbf{W}_\ell^{\text{eff}}$, as described in Section 3.4.4. This comparison is not included to advocate for the use of this approximation; on the contrary, the aim is to demonstrate the merits of using the exact likelihood for Gaussian fields. For this reason, it is not a concern whether appropriate values of ν_{eff} and $\mathbf{W}_\ell^{\text{eff}}$ can be obtained in practice. We simultaneously fitted the three marginals of a $p = 2$ Wishart distribution to the exact marginals, and obtained the following best-fitting values:

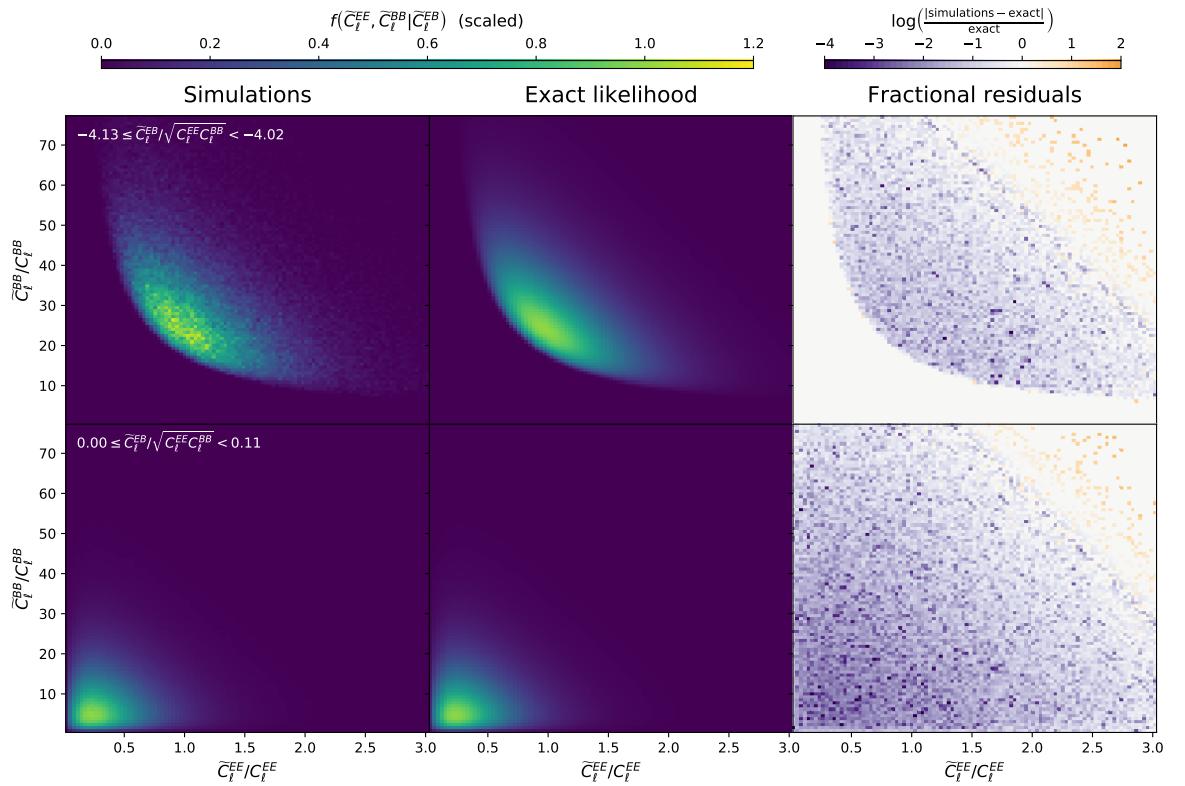


Figure 3.2. Two different slices of the joint $\tilde{C}_\ell^{EE} - \tilde{C}_\ell^{BB}$ distribution for $\ell = 2$, each for one fixed bin of \tilde{C}_ℓ^{EB} . The top row is the slice corresponding to $-4.13 \leq \tilde{C}_\ell^{EB} / \sqrt{\tilde{C}_\ell^{EE} \tilde{C}_\ell^{BB}} < -4.02$ while the bottom row corresponds to $0.00 \leq \tilde{C}_\ell^{EB} / \sqrt{\tilde{C}_\ell^{EE} \tilde{C}_\ell^{BB}} < 0.11$. The left panel in each row is the distribution observed from simulations, while the centre panel is the distribution predicted by the exact likelihood for Gaussian fields. The same colour scale is used for the left and centre panels within each row and has been chosen such that the exact likelihood in each slice runs between 0 and 1. The right panels show logarithmic fractional residuals, as defined in Equation (3.52), and contain only noise due to the finite number of realisations.

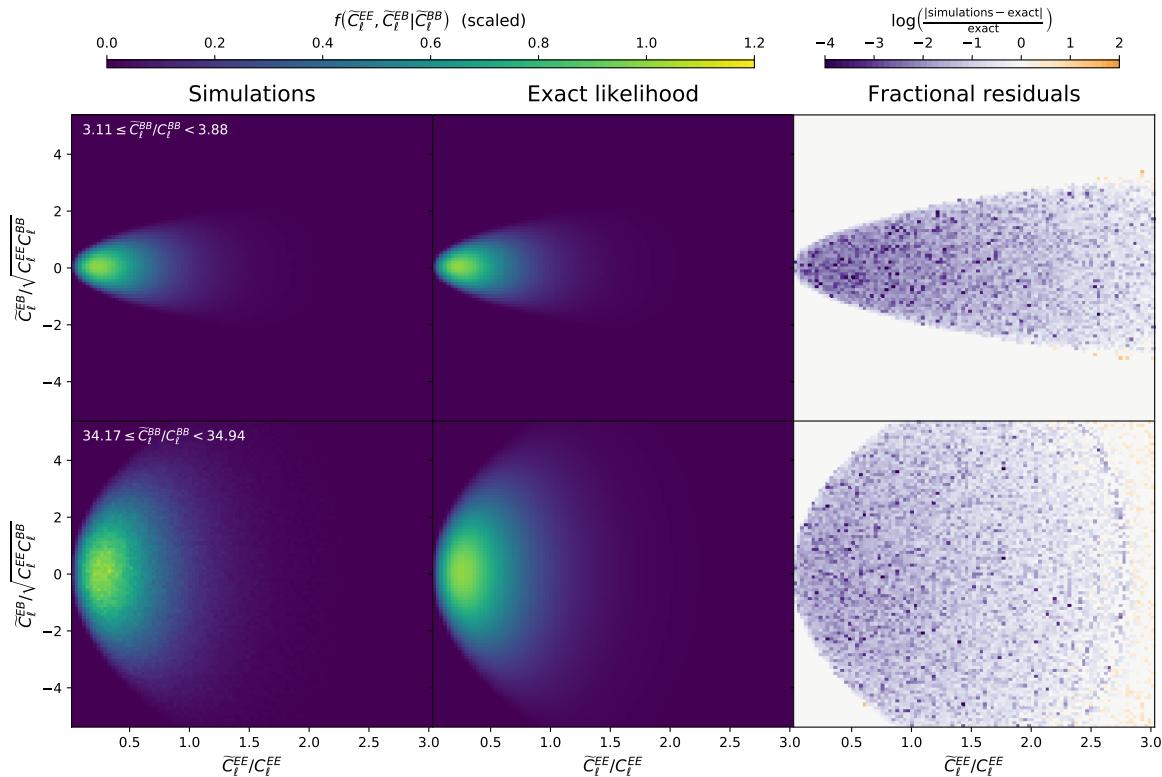


Figure 3.3. As Figure 3.2, but for the $\tilde{C}_\ell^{EE} - \tilde{C}_\ell^{EB}$ distribution at fixed values of \tilde{C}_ℓ^{BB} . The top row corresponds to $3.11 \leq \tilde{C}_\ell^{BB}/C_\ell^{BB} < 3.88$ and the bottom row to $34.17 \leq \tilde{C}_\ell^{BB}/C_\ell^{BB} < 34.94$.

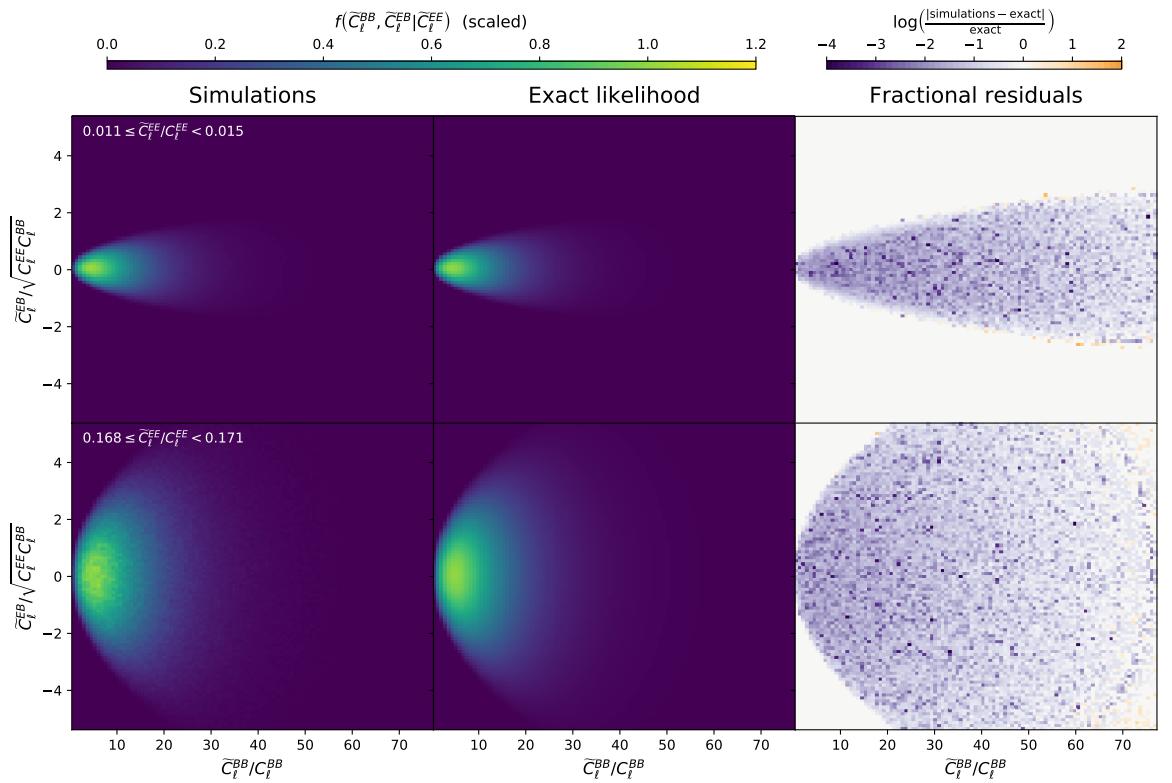


Figure 3.4. As Figure 3.2, but for the $\tilde{C}_\ell^{BB} - \tilde{C}_\ell^{EB}$ distribution at fixed values of \tilde{C}_ℓ^{EE} . The top row corresponds to $0.011 \leq \tilde{C}_\ell^{EE}/C_\ell^{EE} < 0.015$ and the bottom row to $0.168 \leq \tilde{C}_\ell^{EE}/C_\ell^{EE} < 0.171$.

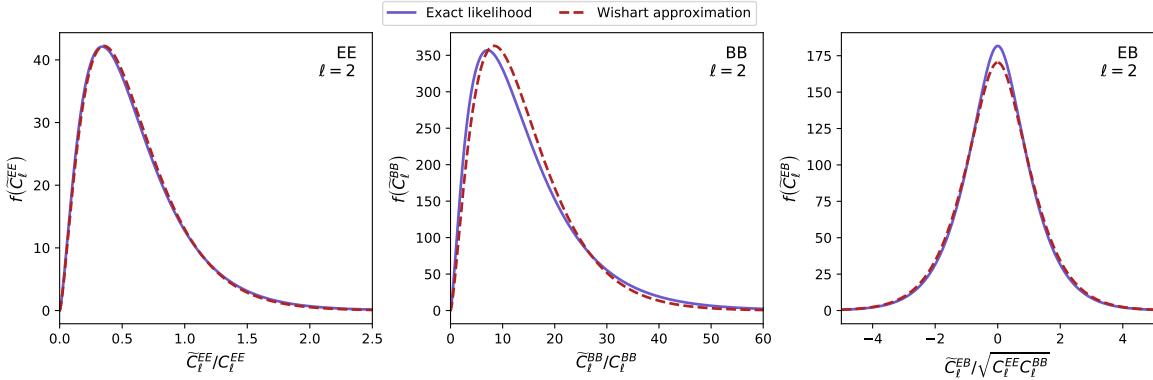


Figure 3.5. Marginal distributions based on the Wishart approximation with best-fitting parameters given in Equation (3.53) (dashed red curves) compared to the exact likelihood for Gaussian fields (solid blue curves).

$$\frac{\nu_{\text{eff}}}{2\ell + 1} = 1.0; \quad \mathbf{W}_\ell^{\text{eff}} = \frac{1}{2\ell + 1} \begin{pmatrix} 0.59C_\ell^{EE} & 0 \\ 0 & 14C_\ell^{BB} \end{pmatrix}. \quad (3.53)$$

There is no reason to expect that these values would also be the best-fitting values for a different ℓ or for a different input cosmology, and they would certainly be different for another mask. The resulting marginal distributions are shown in Figure 3.5, where the simulated histogram is omitted for clarity. The fit is almost perfect for \tilde{C}_ℓ^{EE} , indicating that this marginal distribution closely follows a gamma distribution as in the full-sky case. There is slightly more deviation for \tilde{C}_ℓ^{BB} and \tilde{C}_ℓ^{EB} .

We integrated the Wishart probability density over each histogram bin in three-dimensional space, as with the exact likelihood for Gaussian fields. Figure 3.6 shows one-dimensional slices through this three-dimensional likelihood. For each slice, two dimensions have been fixed at a single histogram bin, and the distribution across the third dimension is shown. While the marginal distributions suggest a near-exact fit for the Wishart approximation, the one-dimensional slices reveal that the approximation incorrectly distributes probability in some parts of the three-dimensional space relative to other parts. The exact likelihood for Gaussian fields, on the other hand, faithfully reproduces the observed distribution throughout. This is also seen in Figure 3.7, which shows three one-dimensional slices in different diagonal directions across the three-dimensional space. In each of these slices, the Wishart approximation overestimates the probability density relative to the true distribution, implying that it must underestimate it in other parts of the distribution, given that the whole distribution is normalised to integrate to 1.

As well as failing to accurately reproduce the full distribution for a fixed ℓ , the Wishart distribution cannot naturally be extended to include correlations between multipoles, whereas the exact likelihood automatically produces the full joint distribution between all multipoles of all power spectra measured on Gaussian fields. In some cases an approxima-

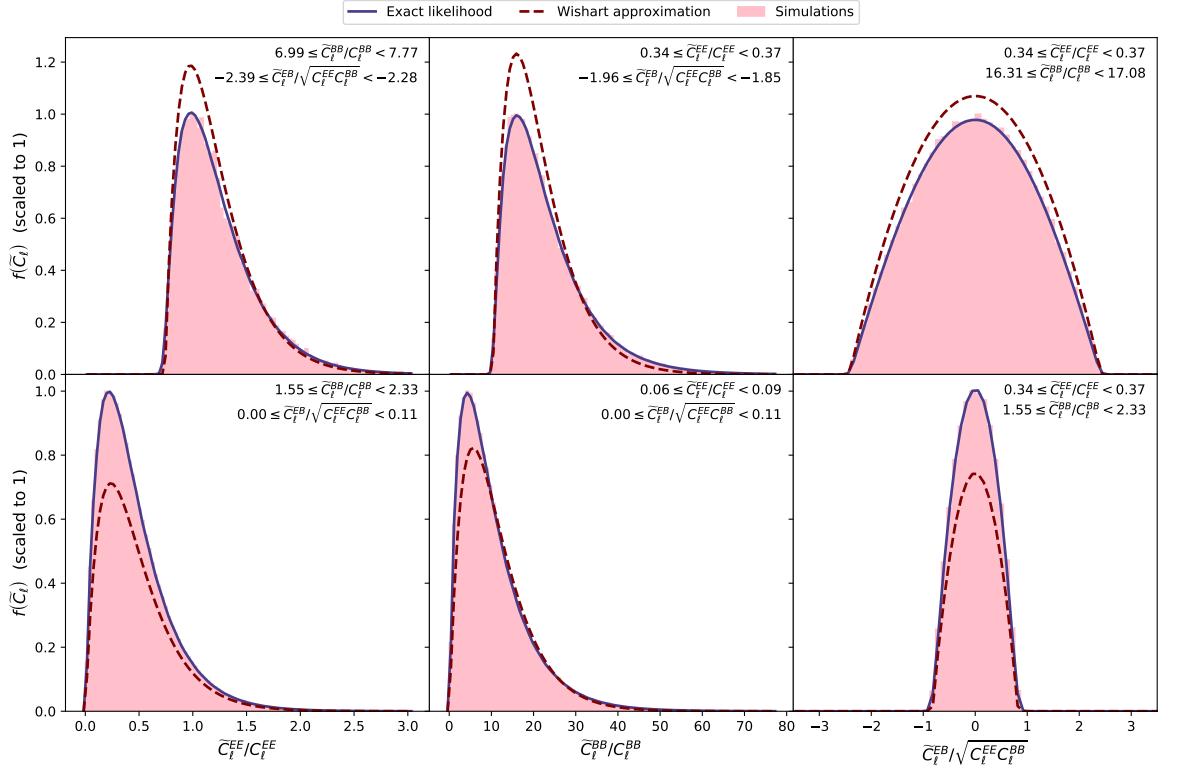


Figure 3.6. One-dimensional slices of the joint distribution of \tilde{C}_ℓ^{EE} , \tilde{C}_ℓ^{BB} and \tilde{C}_ℓ^{EB} for $\ell = 2$. In each slice, two of the values are fixed while the third is allowed to vary.

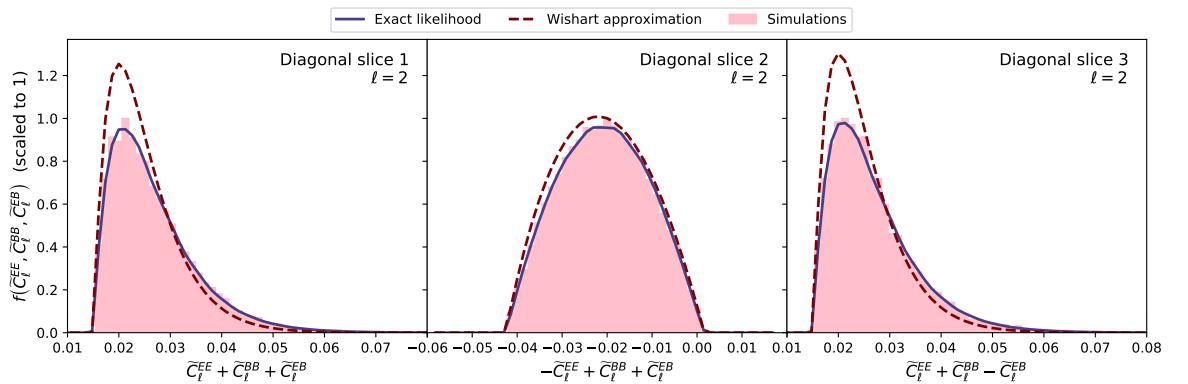


Figure 3.7. As Figure 3.6 but for slices taken in three diagonal directions, each corresponding to a linear combination of the three pseudo- C_ℓ s.

tion may work better over the joint distribution of many multipoles than for a single ℓ (Hamimeche & Lewis 2008), but the inverse can also be true (Elsner & Wandelt 2012). In any case, the choice of approximation for the purposes of this comparison is unimportant: no approximation can completely match the exact distribution.

3.6. Conclusions

This chapter has presented the exact joint likelihood of an arbitrary number of pseudo- C_ℓ estimates from correlated Gaussian fields, valid for both auto- and cross-power spectra and for any mask geometry. The likelihood—given in Equation (3.35)—naturally models both intrinsic correlations between spin-0 and spin-2 fields, and correlations induced by a cut sky which result in the mixing between spherical harmonic coefficients. The pseudo- $a_{\ell m s}$ follow a multivariate Gaussian distribution with covariance matrix elements given by Equations (3.23)–(3.25). The exact joint likelihood for QML power spectrum estimates from Gaussian fields has also been presented, in Equation (3.39). An accurate likelihood function is an essential companion to any estimator for unbiased cosmological inference, but until now a complete likelihood for either the pseudo- C_ℓ or QML estimator on an arbitrary sky has not been known.

Sections 3.4 and 3.5 showed how the exact likelihood can be applied to observations of the polarisation of the Cosmic Microwave Background. This is especially relevant given current and future experiments aimed at detecting primordial B -modes, which require exquisite control of all possible sources of systematic bias. One such source of bias is an inexact likelihood function, so knowledge of the exact likelihood could play an important role in extracting cosmological information from polarisation measurements in an unbiased manner. In particular, it exactly models the leakage of E -mode power into the much smaller B -mode signal. This likelihood also extends naturally to include correlations between temperature anisotropy and polarisation, including cross-correlations between any number of detectors, where each observed temperature or polarisation field may have its own mask. It does not account for weak gravitational lensing of the CMB, which breaks the assumption of Gaussianity at higher multipoles.

The exact likelihood for Gaussian fields could also be extremely useful for weak lensing observations. It will perhaps be most valuable at relatively low multipoles, as this is the regime where the common assumption of a Gaussian likelihood for power spectrum estimates is least applicable due to the considerable skewness in the true likelihood (e.g. Sellentin, Heymans & Harnois-Déraps 2018). These low multipoles correspond to large physical scales, for which it is an excellent approximation to describe the spin-2 cosmic shear field as a Gaussian field. At higher multipoles there will be significant deviations from Gaussianity, so this likelihood cannot be considered exact on small scales. The likelihood naturally extends to describe the full distribution of auto- and cross-power spectra

between an arbitrary number of redshift bins, each with its own mask. It is thus well suited for extracting robust cosmological constraints from tomographic galaxy clustering and weak lensing shear power spectrum measurements in multiple redshift bins, such as those from *Euclid*.

Use of the exact pseudo- C_ℓ likelihood for Gaussian fields is likely to be competitive in terms of speed when considering a small number of bandpower estimates. Its main strength compared to an exact pixel-based likelihood is that for a single power estimate the pseudo- C_ℓ likelihood requires the determinant evaluation of a matrix of size $\sim \ell$ compared to $\sim \ell^2$ for a pixel-based method. This means that it may be evaluated at much higher ℓ than is possible for a pixel-based method. However, once many power estimates are considered the scaling is less competitive. The pixel-based method would offer all additional multipoles $\ell' < \ell$ without significant additional computational cost, while the exact pseudo- C_ℓ likelihood for Gaussian fields scales as $\sim k^N$, where N is the number of (band)power estimates and $k \sim 200$. The cost is driven by the need to evaluate the characteristic function in Equation (3.33) for every value of the vector t . The range of t must cover a wide enough space for the integral in the likelihood expression to converge, while at a sufficiently high resolution so that its curvature is accurately represented. Each point in t -space carries its own determinant or eigenvalue calculation (which is not the case for a single power estimate, due to the simple scaling of the determinant of a single matrix). Use of the exact likelihood for Gaussian fields is therefore only recommended in the case of a very small number of bandpower estimates, and alternative (necessarily approximate) approaches for the joint distribution of many power estimates should be explored. One possibility is the use of a copula with the exact marginal distributions. Copula methods have previously been described in a cosmological context, but only with approximate marginal distributions (Benedek et al. 2009; Sato, Ichiki & Takeuchi 2010; Sato et al. 2011). Alternative approaches previously explored in the literature include approximate extensions to the Wishart distribution to model correlations between multipoles (Hamimeche & Lewis 2008; Mangilli, Plaszczynski & Tristram 2015). Computational limitations in the likelihood calculation may also be mitigated to some extent by potential speed increases at other levels in the inference process such as neural net-assisted sampling (Manrique-Yus & Sellentin 2020).

Despite the limitations of its direct use, knowledge of the exact pseudo- C_ℓ and QML likelihood for Gaussian fields is extremely useful as a starting point and testing benchmark for developing fast, accurate approximations. It is used for this purpose in the work presented in Chapter 4. A common approach to a total likelihood, particularly for CMB observations (e.g. Planck Collaboration et al. 2019) is to use an exact pixel-based likelihood at low multipoles and to switch to an approximate Gaussian power spectrum likelihood for higher multipoles, at the point at which a pixel-based likelihood becomes computationally unfeasible (at $\ell = 29$ in the case of *Planck*, while for weak lensing analyses with many redshift bins an exact pixel-based method may not be feasible at all). Methods derived from this exact likelihood for Gaussian fields may fill an important niche between these two regimes,

allowing the use of an exact or near-exact likelihood up to higher multipoles than is currently possible. This may be a powerful tool for interpreting future observations, given the increased statistical precision that they will offer. This likelihood also has the advantage that it can naturally describe the cross-correlation power spectrum measured between two different maps, in contrast to exact pixel-based methods which are not readily adapted to extracting just the cross-correlation information. Considering only cross-spectra in this way makes the cosmological analysis insensitive to the details of the noise bias, which will be especially relevant for cosmic shear observations, for which shape noise is an important and uncertain factor.

Chapter 4

Sufficiency of a Gaussian likelihood for weak lensing power spectra

4.1. Introduction

Analysis of weak gravitational lensing of distant galaxies by large scale structure is among the most promising methods of constraining theories of dark energy in the near future. As described in Chapter 1, the unprecedented statistical precision offered by such upcoming surveys as *Euclid*, the Rubin Observatory (LSST) and the Square Kilometre Array requires equally unprecedented control of sources of systematic error in order to obtain reliable results. One of the many such sources is the choice of likelihood function, currently routinely assumed to be Gaussian (e.g. Troxel et al. 2018; Hikage et al. 2019; Joachimi et al. 2021).

However, the true likelihood of weak lensing two-point statistics is well known to be non-Gaussian. This has been studied in detail in distributions of simulated data (Sellentin & Heavens 2018; Sellentin, Heymans & Harnois-Déraps 2018; Diaz Rivero & Dvorkin 2020; Louca & Sellentin 2020) and has motivated many derivations of non-Gaussian likelihoods, either approximate or exact under particular conditions (Taruya et al. 2002; Sato, Ichiki & Takeuchi 2010; Sato et al. 2011; Hilbert, Hartlap & Schneider 2011; Keitel & Schneider 2011; Wilking & Schneider 2013; Sellentin 2015; Wilking, Röseler & Schneider 2015; Manrique-Yus & Sellentin 2020; Diaz Rivero & Dvorkin 2020; Hall & Taylor 2022). In particular, in Chapter 3 an exact non-Gaussian likelihood of pseudo- C_ℓ estimates for Gaussian fields was presented.

The impact of wrongly assuming a Gaussian likelihood on cosmological parameter constraints has, however, rarely been investigated in detail. Lin et al. (2020) did so for the shear correlation function in an LSST-like experiment, and found that a Gaussian likelihood is sufficiently accurate for obtaining joint posterior constraints on Ω_m and σ_8 , despite the small 100 deg^2 sky patch used in their tests. This result is in contrast to the earlier work in Hartlap et al. (2009), which found that a Gaussian correlation function likelihood could lead to biased constraints in the same parameters. Taylor et al. (2019) tested the impact of assuming a Gaussian likelihood for the full-sky shear power spectrum on joint constraints of Ω_m and $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$ and found negligible difference in the posterior distribution compared to a likelihood-free approach.

The work presented in this chapter tests the impact of assuming a Gaussian likelihood for a

Euclid-like joint tomographic 3×2 pt power spectrum analysis of weak lensing shear, galaxy clustering and their cross-correlation, on posterior dark energy constraints. The chapter begins with a full-sky setup in Section 4.2, which is extended to a cut sky in Section 4.3 and to non-Gaussian fields in Section 4.4. Conclusions are discussed in Section 4.5.

4.2. Full-sky likelihood

4.2.1. Background

The observable fields considered here are those introduced in Chapter 2: weak lensing shear and galaxy number overdensity. For the majority of this chapter, these fields are treated using Gaussian statistics. This is an approximation, but there are reasons to believe it to be a good one for the purposes of this study, which are discussed in Section 4.4. It is also a necessary starting point, since the only conditions under which the exact joint power spectrum likelihood is both known and tractable is for Gaussian fields on the full sky. Therefore, results will first be obtained for Gaussian fields. Section 4.4 argues that these results hold for real observable fields, and also contains an analysis of the distribution of power spectrum estimates from N-body simulations.

4.2.1.1. Wishart distribution

For correlated Gaussian fields observed on the full sky, the set of observed C_ℓ s follows a Wishart distribution, independently for each ℓ (see Percival & Brown 2006 for a derivation in the case of CMB temperature and polarisation). This distribution can be parametrised using the degrees of freedom ν and $p \times p$ scale matrix \mathbf{V} , in which case the probability distribution function (PDF) for random matrix \mathbf{X} is

$$f_{\mathcal{W}}(\mathbf{X}|\nu, \mathbf{V}) = \frac{|\mathbf{X}|^{(\nu-p-1)/2} \exp[-\text{trace}(\mathbf{V}^{-1}\mathbf{X})/2]}{2^{\nu p/2} |\mathbf{X}|^{\nu/2} \Gamma_p(\nu/2)}, \quad (4.1)$$

where Γ_p is the multivariate gamma function. For an N -bin tomographic 3×2 pt analysis, the set of observed C_ℓ s can be written as a $2N \times 2N$ symmetric matrix, $\widehat{\mathbf{C}}_\ell$:

$$\widehat{\mathbf{C}}_\ell = \begin{pmatrix} \widehat{C}_\ell^{n(1)n(1)} & \widehat{C}_\ell^{n(1)E(1)} & \dots & \widehat{C}_\ell^{n(1)n(N)} & \widehat{C}_\ell^{n(1)E(N)} \\ \widehat{C}_\ell^{n(1)E(1)} & \widehat{C}_\ell^{E(1)E(1)} & \dots & \widehat{C}_\ell^{E(1)n(N)} & \widehat{C}_\ell^{E(1)E(N)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{C}_\ell^{n(1)n(N)} & \widehat{C}_\ell^{E(1)n(N)} & \dots & \widehat{C}_\ell^{n(N)n(N)} & \widehat{C}_\ell^{n(N)E(N)} \\ \widehat{C}_\ell^{n(1)E(N)} & \widehat{C}_\ell^{E(1)E(N)} & \dots & \widehat{C}_\ell^{n(N)E(N)} & \widehat{C}_\ell^{E(N)E(N)} \end{pmatrix}, \quad (4.2)$$

where n represents the number overdensity field and E the shear E -mode, and $\widehat{C}_\ell^{X(i)Y(j)}$ is the observed cross-power between redshift bins i and j . For Gaussian fields, $\widehat{\mathbf{C}}_\ell$ follows a

Wishart distribution with parameters

$$\widehat{\mathbf{C}}_\ell \sim \mathcal{W} \left(\nu = 2\ell + 1, \mathbf{V} = \frac{\mathbf{C}_\ell}{2\ell + 1} \right), \quad (4.3)$$

where \mathbf{C}_ℓ is the symmetric positive definite matrix of underlying C_ℓ s analogous to $\widehat{\mathbf{C}}_\ell$,

$$\mathbf{C}_\ell = \begin{pmatrix} C_\ell^{n(1)n(1)} & C_\ell^{n(1)E(1)} & \dots & C_\ell^{n(1)n(N)} & C_\ell^{n(1)E(N)} \\ C_\ell^{n(1)E(1)} & C_\ell^{E(1)E(1)} & \dots & C_\ell^{E(1)n(N)} & C_\ell^{E(1)E(N)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_\ell^{n(1)n(N)} & C_\ell^{E(1)n(N)} & \dots & C_\ell^{n(N)n(N)} & C_\ell^{n(N)E(N)} \\ C_\ell^{n(1)E(N)} & C_\ell^{E(1)E(N)} & \dots & C_\ell^{n(N)E(N)} & C_\ell^{E(N)E(N)} \end{pmatrix}. \quad (4.4)$$

The order of rows and columns in \mathbf{C}_ℓ and $\widehat{\mathbf{C}}_\ell$ is arbitrary, provided it is consistent between the two matrices. For simplicity shape noise has been ignored in Equation (4.2) and Equation (4.4), but this may be included by replacing each C_ℓ in the diagonal with $C_\ell + N_\ell$, where N_ℓ is the corresponding noise power. Noise is included in the *Euclid*-like setup described in Section 4.2.2. This setup may also be trivially extended to include a shear B -mode.

It follows that the exact likelihood for a set of observed power spectra from correlated Gaussian fields on the full sky is a product of Wishart distributions, one for each ℓ , each following Equation (4.3).

4.2.1.2. Gaussian distribution

As introduced in Chapter 2, the multivariate Gaussian distribution, parametrised by mean vector μ and covariance matrix Σ , for length- k random vector x has PDF

$$f_{\mathcal{N}}(x|\mu, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right]. \quad (4.5)$$

We may define a vector of observed C_ℓ s containing the unique elements of the matrix $\widehat{\mathbf{C}}_\ell$. If $\widehat{\mathbf{C}}_\ell$ obeys Equation (4.3), then the expectation value of this vector will be the corresponding elements of \mathbf{C}_ℓ ; i.e., the expectation value of any given observed \widehat{C}_ℓ is the corresponding underlying C_ℓ . The covariance matrix of this vector has elements given by the well-known general expression for the covariance of full-sky C_ℓ estimates,

$$\text{Cov} \left(\widehat{C}_\ell^{\alpha\beta}, \widehat{C}_{\ell'}^{\gamma\epsilon} \right) = \frac{\delta_{\ell\ell'}}{2\ell + 1} \left(C_\ell^{\alpha\gamma} C_\ell^{\beta\epsilon} + C_\ell^{\alpha\epsilon} C_\ell^{\beta\gamma} \right), \quad (4.6)$$

where δ is the Kronecker delta. Therefore, the exact distribution of full-sky power estimates (Equation 4.3) may be approximated by a Gaussian distribution having the same mean and covariance.

It turns out that this approximation performs much better if the covariance is fixed at some fiducial cosmology, rather than being re-evaluated at each set of theory C_ℓ s being considered in a likelihood analysis. This is explored in some detail in Hamimeche & Lewis (2008) and Carron (2013), where it is also shown that allowing the covariance to vary as a function of cosmology violates the Cramér-Rao bound. This is also discussed in the methodology paper of the recent KiDS-1000 analysis (Joachimi et al. 2021). Therefore, this is the approximation that is tested in this chapter: the term ‘Gaussian likelihood’ should be taken to refer to the version of Equation (4.5) where Σ is fixed at some fiducial cosmology. The impact of the choice of fiducial cosmology is explored in Section 4.2.4.2.

As will be discussed in more detail in Section 4.3.1, the marginal distributions of a Gaussian-distributed vector have zero skewness and excess kurtosis, which is not the case for the Wishart distribution. Since here the mean and variance of the Gaussian distribution are fixed to be equal to those of the Wishart distribution, the inaccuracy of the Gaussian likelihood approximation in describing the true marginal distributions will be largely captured by the skewness and excess kurtosis. However, for the Wishart distribution both of these quantities decrease as power laws in $2\ell + 1$, and the behaviour of the cut-sky likelihood is similar. Therefore, the inaccuracy of the Gaussian likelihood will be most pronounced for low ℓ , corresponding to large physical scales. This is described in more detail in Section 4.3.1.

4.2.2. Full sky: Methodology

The tests in this section involve comparing exact posterior distributions, obtained with the Wishart likelihood, to approximate posterior distributions obtained with the Gaussian likelihood. The mean, maximum and standard deviation of single-parameter posteriors are studied in Section 4.2.3, and the contours of two-dimensional posteriors in Section 4.2.4. As introduced in Chapter 2, the posterior distribution of model parameters θ from observed data d , $p(\theta | d)$ is calculated by evaluating Bayes’ theorem,

$$p(\theta | d) \propto \pi(\theta) f(d | \theta). \quad (4.7)$$

The normalisation constant is formally given by the Bayesian evidence, but in this work a manual normalisation is used assuming a uniform prior $\pi(\theta)$, chosen to be sufficiently broad as to negligibly affect the posterior distribution. The remaining three ingredients are the model predictions, which are deterministic functions of θ , the (mock) observation d , and the likelihood function $f(d | \theta)$ which connects them. Each of these are described below.

4.2.2.1. Theory

For this work, regular grids of one, two or three cosmological parameters from (w_0 , w_a , Ω_m) were used, with all other parameters held to a fixed value. These grids were generated using CosmoSIS (Zuntz et al. 2015). The pipeline consisted of the following CosmoSIS standard library modules:

1. CAMB version Jan15, to calculate the linear matter power spectrum (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012);
2. Halofit_Takahashi version Camb-Nov-13, to compute the non-linear matter power spectrum (Smith et al. 2003, Takahashi et al. 2012; CosmoSIS module by A. Lewis & S. Bird);
3. no_bias version 1, to calculate the galaxy power spectrum with no galaxy bias—this choice was made for simplicity, since galaxy bias is irrelevant to the tests presented here;
4. gaussian_window version 1, to calculate Gaussian redshift distributions—for this work, 5 bins centred on $z = 0.65, 0.95, 1.25, 1.55, 1.85$ were used, each with $\sigma = 0.3$;
5. project_2d version 1.0, to calculate projected galaxy and shear power spectra applying the Limber approximation—the accuracy of the Limber approximation is also irrelevant for the purposes of the tests presented here. A small modification was applied to this module to output linearly spaced C_ℓ s for the full multipole range ($2 \leq \ell \leq 2000$).

4.2.2.2. Mock observations

To generate mock observations, output power spectra were taken from the pipeline described above, with zero shear B -mode signal. A noise contribution, N_ℓ , was then added to each auto-power spectrum (see Chapter 2):

$$N_\ell^{n(i)n(i)} = \frac{1}{N_i}; \quad (4.8)$$

$$N_\ell^{E(i)E(i)} = N_\ell^{B(i)B(i)} = \frac{\sigma_\varepsilon^2}{N_i}, \quad (4.9)$$

where N_i is the galaxy number density per redshift bin and σ_ε is the intrinsic ellipticity dispersion per component. For this work, a *Euclid*-like number density of $30/\text{arcmin}^2$, split equally between the five redshift bins, and a value of $\sigma_\varepsilon = 0.3$, were used.

This results in 120 input power spectra, of which 65 relate to shear B -mode so are zero or noise-only. From these, the healpy Python implementation of the HEALPix software (Górski et al. 2005; Zonca et al. 2019) was used to generate 15 correlated maps, three per

redshift bin, having the full set of $3 \times 2\text{pt}$ correlations, including the noise contribution and a proper spin-2 treatment of shear. The default resolution used in these tests is $\text{nside} = 1024$ (corresponding to an angular scale of 3.4 arcmin) and $\ell_{\max} = 2000$; any departure from this is noted in the text. `healpy` was used again to measure the 120 observed full-sky power spectra from these maps.

No angular binning to form bandpowers is used here, since the exact likelihood in this case would no longer be a Wishart distribution. Instead, it would follow a more complicated distribution, whose PDF could in principle be obtained either as a convolution of Wishart PDFs or from the general PDF of quadratic forms in Gaussian variables, analogous to the exact pseudo- C_ℓ likelihood derived in Chapter 3. The feasibility of such an approach in practice is unclear, and is not the focus of this work. Furthermore, the distribution of individual C_ℓ estimates should be more non-Gaussian than the distribution of bandpowers, since each bandpower has more contributing modes. This implies that the results obtained here should be taken in this regard as a lower limit on the accuracy of the Gaussian likelihood.

4.2.2.3. Likelihoods

B -mode power spectra are excluded from the likelihood analysis, leaving 55 power spectra as input to the likelihoods. Custom code was implemented in Python to evaluate each log-likelihood at every grid point. For the Wishart likelihood, the SciPy¹ Wishart log-PDF function (Virtanen et al. 2020) was used, which implements Equation (4.3). For the Gaussian likelihood, a custom implementation of the Gaussian log-PDF in Equation (4.5) with precomputed inverse covariance was used, neglecting the determinant term since it is constant when the covariance is fixed. Each log-likelihood is exponentiated and normalised separately.

4.2.3. Full sky: Summary statistics

For the tests in this subsection, 7 000 mock observations were generated following the steps outlined above (Section 4.2.2.2), but with $\ell_{\max} = 100$ to keep computation time and data volume within reasonable limits. This isolates the part of the data vector for which the Gaussian likelihood should be expected to perform worst, since the likelihood is most non-Gaussian at low ℓ (see Section 4.3). For each realisation, a single-parameter likelihood analysis on w_0 was run with other parameters fixed. The sections below study the distributions of the maximum, mean and standard deviation of the resulting one-dimensional posterior distribution across all realisations.

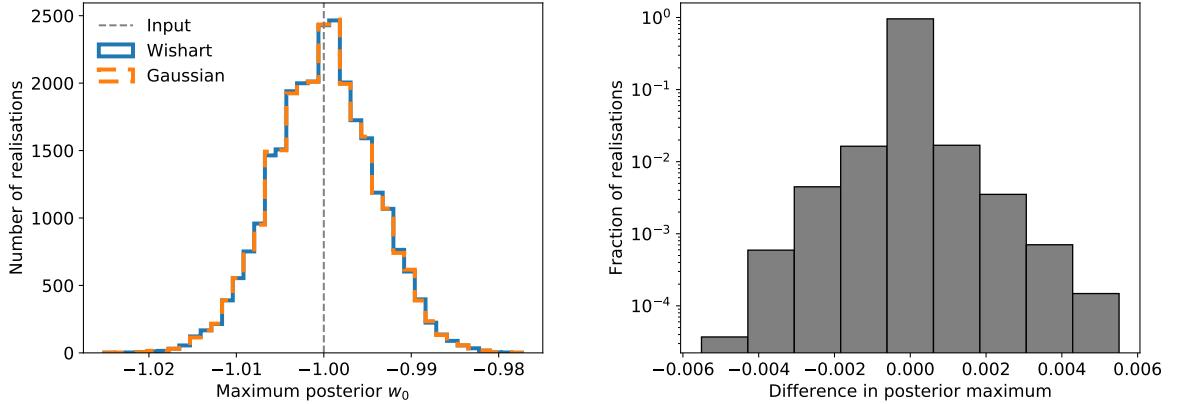


Figure 4.1. *Left:* Distribution of posterior maxima returned by the Gaussian likelihood compared to the true, Wishart likelihood. *Right:* Per-realisation difference between the posterior maximum returned by the two likelihoods, with a positive difference representing a higher value of w_0 for the Gaussian likelihood.

4.2.3.1. Posterior maximum

For a single spin-0 Gaussian field, the Gaussian likelihood with fixed variance is guaranteed to give the same posterior maximum as the true likelihood, for a flat prior (Carron 2013). Hamimeche & Lewis (2008) investigated the extension to correlated fields and found that while the exactness of this relation does not hold, the Gaussian likelihood will still return the correct posterior maximum as long as the fiducial model is proportional to the model which maximises the likelihood. It is argued in that paper that for models which vary smoothly with ℓ , this will often hold approximately even if it does not hold exactly.

The left panel of Figure 4.1 shows the distribution of posterior maxima obtained from the two likelihoods for the 27 000 realisations. The distributions are almost indistinguishable. The right panel shows the per-realisation difference between the posterior maximum returned by the two likelihoods. The Gaussian returns the correct maximum for 95.7 per cent of the realisations, and for the remainder it is wrong by no more than four grid points, which have a size of $\Delta w_0 = 1.25 \times 10^{-3}$. This demonstrates that—as predicted in Hamimeche & Lewis (2008)—the maximum-posterior property of the Gaussian likelihood holds to a very good approximation in practice for correlated fields.

4.2.3.2. Posterior mean and standard deviation

Along with the posterior maximum, two other summary statistics for which it is perhaps most important for an approximate likelihood to return accurate values are the posterior mean and standard deviation. Unlike the posterior maximum, there is no general property of the Gaussian likelihood which says that it should return approximately correct values of these quantities. However, this appears to be the case on average: Figure 4.2 shows the

¹ <https://scipy.org>

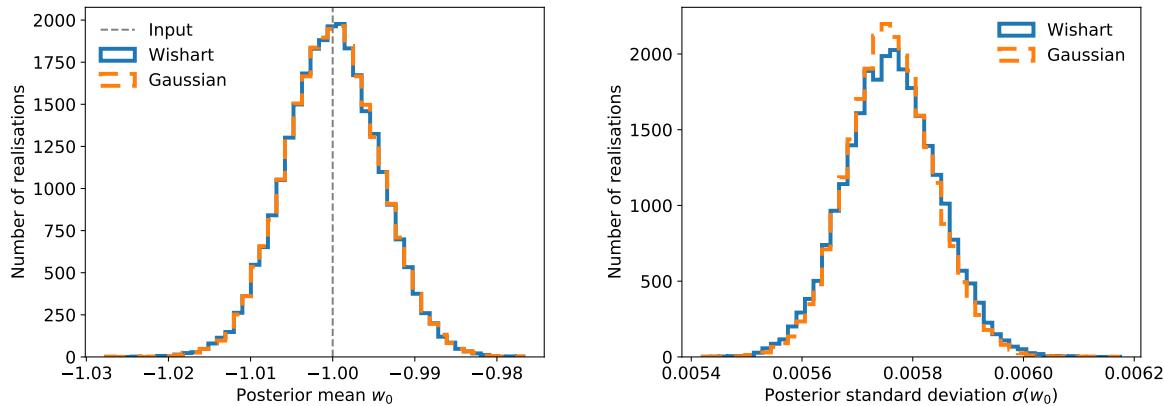


Figure 4.2. *Left:* Distribution of posterior means returned by the Gaussian likelihood compared to the true, Wishart likelihood. *Right:* Distribution of posterior standard deviations for the two likelihoods.

distribution of posterior means (left panel) and standard deviations (right panel) for the Gaussian likelihood compared to the true, Wishart likelihood. The distributions of means are almost indistinguishable. The distributions of standard deviations are very similar, though there is some visible discrepancy. Further investigation revealed that the Gaussian overestimates the standard deviation on realisations for which the true standard deviation is low (relative to its average over all realisations) and underestimates the standard deviation on realisations for which the true standard deviation is high. This effect has a magnitude of order 1 per cent of the true standard deviation. This is highly likely to be an acceptable level of inaccuracy, and is also expected to be smaller still when ℓ_{\max} is higher than the value of 100 used here.

4.2.4. Full sky: Posterior contours

As described in Chapter 2, cosmological parameter constraints are often visualised using two-dimensional contour plots, with the contours representing particular credible regions. Here, the accuracy of the Gaussian likelihood is tested in this regard. A single mock observation is used, produced following the method described in Section 4.2.2.2 with $\ell_{\max} = 2000$. This realisation was produced at random, but the results presented here have been checked with different realisations, and all give identical results in terms of level of agreement between the two likelihoods. As is the case in a real experiment, the posterior constraints are not centred on the input cosmology, due to the sizeable contribution from cosmic variance inherent in a single realisation.

In most cases, a two-parameter likelihood analysis in (w_0, w_a) is performed to keep computational costs down, but a three-parameter example is also provided to demonstrate that marginalisation over a third parameter does not affect the level of agreement between likelihoods. All two-dimensional posteriors are presented in terms of $1\text{--}3\sigma$ contours, using the

shorthand convention (see Chapter 2) that 1, 2 and 3σ represent 68.3, 95.4 and 99.7 per cent posterior probability.

4.2.4.1. Baseline setup

The baseline full-sky test setup used here is as follows. The sensitivity of the results presented in this chapter to the details of this setup are tested in Section 4.2.4.2.

1. Five redshift bins (see Section 4.2.2.1), each with galaxy number overdensity and shear E -mode fields;
2. All 55×2 pt power spectra between these ten fields; i.e., galaxy–galaxy, shear–shear and galaxy–shear;
3. Gaussian noise assuming a *Euclid*-like number density evenly split between bins (see Section 4.2.2.2);
4. Multipole range $2 \leq \ell \leq 2000$;
5. Fiducial cosmology for Gaussian covariance equal to the true input cosmology used to generate the mock observation.

Figure 4.3 shows two- and one-dimensional marginalised posterior distributions obtained from a three-parameter likelihood analysis with the Wishart and Gaussian likelihoods. The results from the two likelihoods are visually indistinguishable, showing that under the baseline setup the Gaussian likelihood is sufficiently accurate.

4.2.4.2. Robustness to deviation from baseline setup

It is important to check that the impressive degree of accordance between the Wishart and Gaussian likelihoods in Figure 4.3 is not a result of any specific choices made in the baseline setup outlined above. The robustness of these results to deviations from this baseline setup is now tested. For these tests, a two-parameter likelihood analysis was performed with other parameters fixed.

Fiducial cosmology

The Gaussian likelihood with fixed covariance requires choosing a fiducial cosmology at which to evaluate the covariance. In the baseline setup, the fiducial cosmology was chosen to be equal to the true input cosmology that was used to generate the mock observation. In a real analysis this would not be possible, since the true cosmology is unknown. To model this effect, the analysis was repeated with the fiducial cosmology chosen to be distant from the true cosmology. Figure 4.4 shows one example, for which the fiducial cosmology is excluded at more than 10σ and yet this does not appear to affect the accuracy of the Gaussian likelihood. Further tests showed that the accuracy does eventually diminish, but

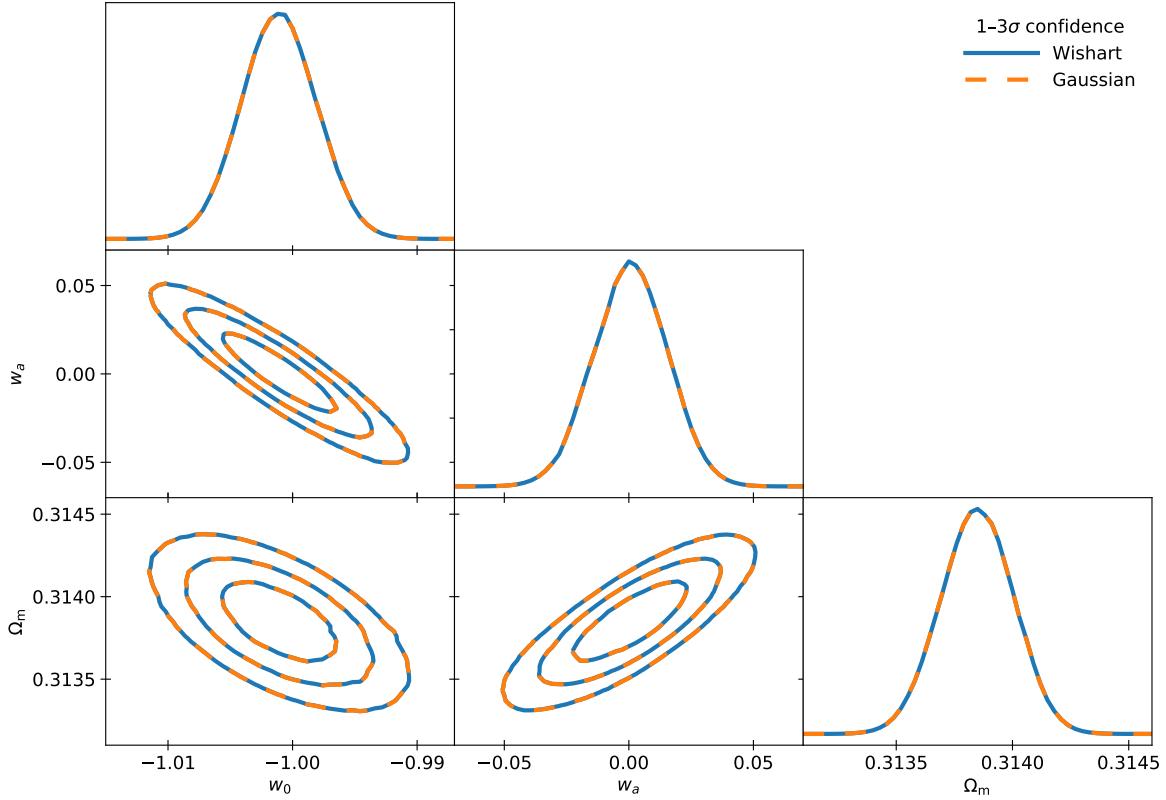


Figure 4.3. Two- and one-dimensional marginalised posteriors from a three-parameter likelihood analysis using the Gaussian likelihood compared to the true, Wishart likelihood.

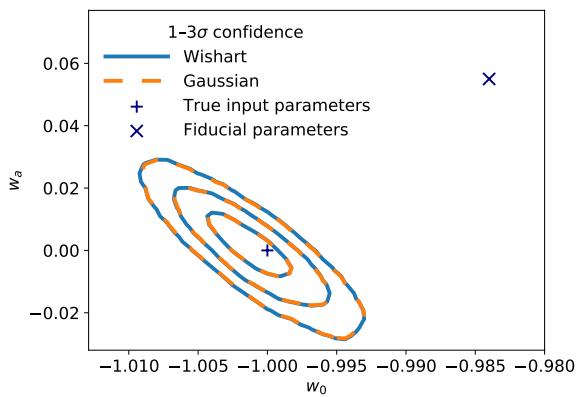


Figure 4.4. Posterior distribution of w_0 and w_a with other parameters fixed, where the fiducial cosmology used to evaluate the Gaussian covariance is excluded at high confidence.

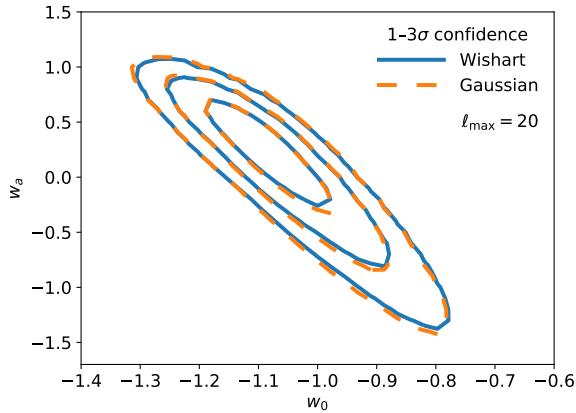


Figure 4.5. Posterior distribution of w_0 and w_a with other parameters fixed, with only $\ell = 2\text{--}20$ included in the likelihood.

only when the fiducial cosmology and the true cosmology are unrealistically far apart (e.g. using a fiducial $w_0 = -0.2$ and a true $w_0 = -1.0$). Even in this case, it is the size and shape of the posterior distribution that is affected, much more than its location. In any real analysis, if the fiducial cosmology were excluded at high confidence then the analysis should be repeated with a fiducial cosmology consistent with the data. Therefore, even if posterior parameter constraints in the initial case were inaccurate due to the choice of fiducial cosmology, they would converge onto the correct constraints through this process.

ℓ range

The Gaussian likelihood should be expected to perform best at high ℓ , as the true likelihood gradually tends to Gaussian by the central limit theorem as more $a_{\ell m}$ s contribute to each C_ℓ estimate (see Section 4.3). Therefore, as ℓ_{\max} is reduced for a constant ℓ_{\min} , the accuracy of the Gaussian likelihood should decrease. This expected behaviour is observed, but it is surprisingly weak. Figure 4.5 shows the posterior distribution obtained with $\ell_{\max} = 20$. While there is some disagreement between the two sets of contours, the Gaussian likelihood is still very clearly able to recover the non-Gaussian shape of the true posterior.

Noise

Even under the assumption of Gaussian fields and Gaussian noise, the level of noise has a theoretical impact on the accuracy of the Gaussian likelihood. This is because the noise power spectrum is flat, while the signal power spectra all decrease with ℓ . For any noise level, there is a threshold ℓ above which the noise dominates the signal. Increasing the noise level decreases this threshold, meaning that a greater fraction of the overall constraining power of the data comes from lower ℓ . Therefore, increasing the noise level relatively upweights the contribution of lower ℓ , which—as discussed above—is the subset of the data for which the Gaussian likelihood should perform worst. However, this does not

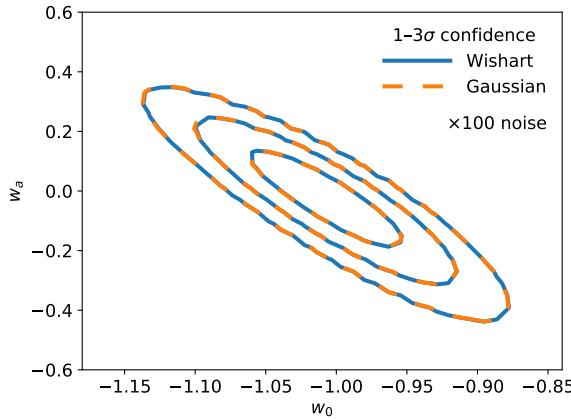


Figure 4.6. Posterior distribution of w_0 and w_a with other parameters fixed, with noise at $100 \times$ *Euclid*-like levels.

appear to have a noticeable effect for realistic noise levels. Figure 4.6 shows the posterior distributions obtained with $100 \times$ *Euclid*-like noise, achieved by assuming a total galaxy number density of $0.3/\text{arcmin}^2$. The effect of decreasing noise was also tested (and even switching it off entirely for a reduced two-redshift-bin setup), though for Gaussian fields and Gaussian noise this should not decrease the accuracy of the Gaussian likelihood; rather, it should increase following the inverse of the above argument. In both cases this did not lead to any visible discrepancy between the two posteriors.

Other aspects of the baseline setup were also varied, including testing with a single power spectrum and testing other parameter combinations ($\Omega_m - \sigma_8$, $w_0 - n_s$, $\Omega_m - \sigma_8 - w_0$) but none of these made any visible difference to the level of accordance between the two likelihoods. One change that did make a significant difference was allowing the covariance matrix in the Gaussian likelihood to vary across parameter space rather than being fixed, confirming that this aspect is crucial to the accuracy of the Gaussian likelihood. With the covariance fixed, we can conclude that the Gaussian likelihood is sufficiently accurate for full-sky power spectra from Gaussian fields.

4.3. Cut-sky likelihood

It cannot be necessarily assumed that the accuracy of the Gaussian likelihood on the full sky will extend to the cut sky, where the situation is more complicated. Although the exact cut-sky likelihood under the assumption of Gaussian fields is known (Chapter 3), it is only feasible to use in its exact form in specific low-dimensional cases. On the full sky, all $a_{\ell m}$ s are independent, and are identically distributed for a given ℓ . The introduction of a mask mixes the $a_{\ell m}$ s (Lewis, Challinor & Turok 2001; Brown, Castro & Taylor 2005), breaking these two properties. For an exact treatment, it becomes necessary to keep track of the relationship between all $a_{\ell m}$ s, which quickly becomes impossible for a high-dimensional

analysis such as this one.

This motivates an alternative approach, which is taken here: after having tested the accuracy of the Gaussian likelihood on the full sky—where the exact likelihood is both known and tractable—the ways in which a sky cut might decrease this accuracy are now carefully considered. To do this, the non-Gaussianity of the cut-sky likelihood is compared to the non-Gaussianity of the full-sky likelihood. We can utilise Sklar’s theorem, which states that any multivariate probability distribution may be separated into its marginal distributions and its dependence structure (Sklar 1959). The dependence structure is called the copula, though for clarity this term is not used here to avoid confusion with the method of forming an approximate joint distribution by combining separate approximations for marginals and the copula, commonly using a Gaussian copula, which was mentioned in Chapter 3 (see Benabed et al. 2009; Sato, Ichiki & Takeuchi 2010; Sato et al. 2011, for more discussion of this method in a cosmological context). The non-Gaussianity of the marginal distributions is therefore considered in Section 4.3.1, and the non-Gaussianity of the dependence structure in Section 4.3.2, in each case comparing between the full-sky and cut-sky likelihoods.

4.3.1. Cut sky: Effect on marginal distributions

The investigation into the non-Gaussianity of marginal distributions focuses on auto-spectra, which by their positive-definite nature are much more non-Gaussian than cross-spectra on both the full and cut sky (Percival & Brown 2006; Chapter 3). Non-Gaussianity is here quantified using the skewness and excess kurtosis, following Lin et al. (2020) and Diaz Rivero & Dvorkin (2020), since both vanish for a Gaussian distribution. These can be written in terms of the mean $E[X]$ and standard deviation $\text{Std}(X)$ of a random variable X as

$$\text{Skew}(X) = E \left[\left(\frac{X - E[X]}{\text{Std}(X)} \right)^3 \right]; \quad (4.10)$$

$$\text{Ex kurt}(X) = E \left[\left(\frac{X - E[X]}{\text{Std}(X)} \right)^4 \right] - 3. \quad (4.11)$$

The marginal distribution of a full-sky auto- C_ℓ (the diagonal elements of Equation 4.2) is a gamma distribution, which under the (k, θ) parametrisation has PDF

$$f_\Gamma(x|k, \theta) = \frac{x^{k-1} \exp[-x/\theta]}{\Gamma(k)\theta^k}, \quad (4.12)$$

where Γ is the gamma function. This distribution has skewness and excess kurtosis

$$\text{Skew}(X) = \frac{2}{\sqrt{k}}; \quad (4.13)$$

$$\text{Ex kurt}(X) = \frac{6}{k}. \quad (4.14)$$

The full-sky likelihood corresponds to parameter values (Percival & Brown 2006; Hamimeche & Lewis 2008; Sellentin, Heymans & Harnois-Déraps 2018)

$$\hat{C}_\ell \sim \Gamma \left(k = \frac{2\ell + 1}{2}, \theta = \frac{2C_\ell}{2\ell + 1} \right). \quad (4.15)$$

Both the skewness and kurtosis therefore depend only on ℓ , and are both power laws in $2\ell + 1$:

$$\text{Skew}(\hat{C}_\ell) = \sqrt{8} [2\ell + 1]^{-1/2}; \quad (4.16)$$

$$\text{Ex kurt}(\hat{C}_\ell) = 12 [2\ell + 1]^{-1}. \quad (4.17)$$

The skewness and excess kurtosis of the cut-sky likelihood may be derived from the pseudo- C_ℓ marginal characteristic function (CF),

$$\varphi_{\tilde{C}_\ell}(t) = \prod_j (1 - 2i\lambda_j t)^{-1/2}, \quad (4.18)$$

where $\{\lambda_j\}$ are the eigenvalues of $\mathbf{M}\Sigma$, the product of the pseudo- $a_{\ell m}$ covariance matrix Σ with \mathbf{M} , the selection matrix picking out the relevant elements of Σ for the C_ℓ in question (see Chapter 3). Equation (4.18) may be identified as a product of gamma distribution CFs:

$$\varphi_\Gamma(t) = (1 - \theta it)^{-k}, \quad (4.19)$$

each with parameters $k = 1/2$, $\theta = 2\lambda_j$. Since the CF of a sum of independent random variables is equal to the product of the individual CFs, it follows that the marginal distribution of a pseudo- C_ℓ estimate is identical to that of a sum of independent gamma-distributed variables. This allows the calculation of the cut-sky skewness and excess kurtosis in terms of the eigenvalues λ_j of $\mathbf{M}\Sigma$:

$$\text{Skew}(\tilde{C}_\ell) = \frac{2^{3/2} \sum_j \lambda_j^3}{\left[\sum_j \lambda_j^2 \right]^{3/2}}; \quad (4.20)$$

$$\text{Ex kurt}(\tilde{C}_\ell) = \frac{12 \sum_j \lambda_j^4}{\left[\sum_j \lambda_j^2 \right]^2}. \quad (4.21)$$

Figure 4.7 shows the full- and cut-sky skewness as a function of ℓ , up to $\ell = 80$, for a *Euclid*-

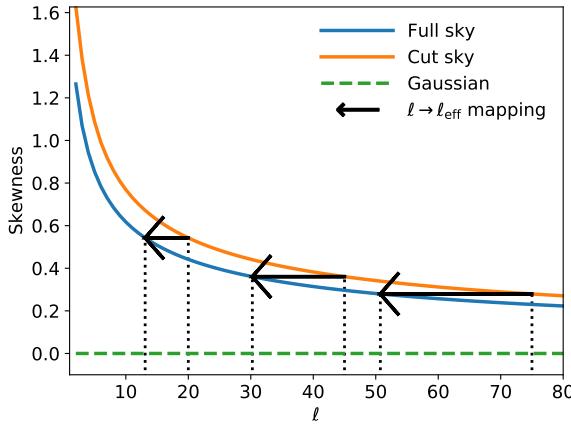


Figure 4.7. Skewness of the full-sky and cut-sky marginal auto- C_ℓ distribution as a function of ℓ , where the cut-sky result is for a *Euclid*-like mask. The arrows demonstrate the $\ell \rightarrow \ell_{\text{eff}}$ mapping described in Section 4.3.1.1.

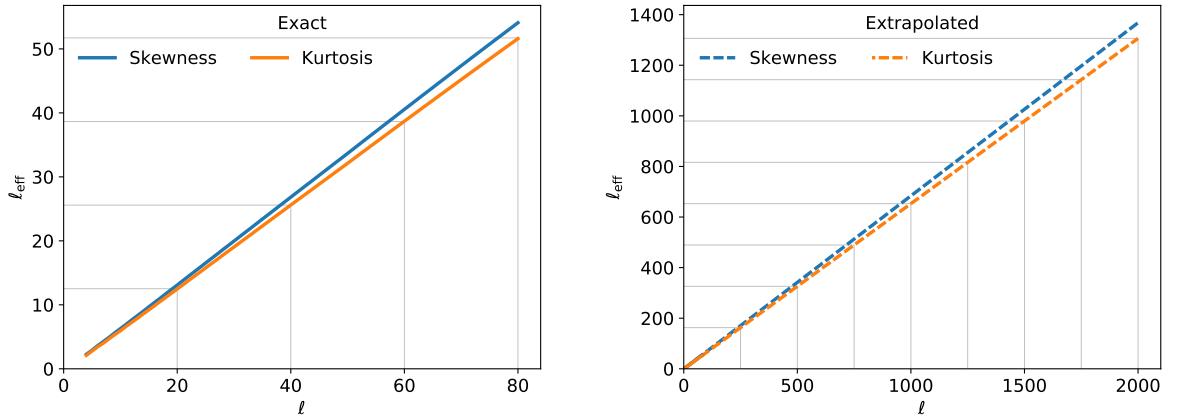


Figure 4.8. *Left:* $\ell \rightarrow \ell_{\text{eff}}$ mapping derived from equating the skewness (blue) and excess kurtosis (orange) of the full-sky and cut-sky likelihoods. *Right:* Extrapolated to $\ell = 2000$.

like mask incorporating the survey footprint and a bright star mask ($f_{\text{sky}} = 30.7$ per cent). Both curves are smoothly decreasing, with the cut-sky skewness systematically higher. The kurtosis exhibits a similar behaviour. The marginal distributions of the likelihood, therefore, are more non-Gaussian on the cut sky than on the full sky. The impact of this additional non-Gaussianity is now investigated.

4.3.1.1. Impact of additional non-Gaussianity

We can take advantage of the fact that both skewness and kurtosis are higher on the cut sky and that both decrease smoothly with ℓ to define an ‘effective ℓ ’, ℓ_{eff} , for each ℓ by equating the full- and cut-sky skewness, and the same for kurtosis. This process is demonstrated by the arrows in Figure 4.7. This $\ell \rightarrow \ell_{\text{eff}}$ mapping is shown in the left panel of Figure 4.8, and turns out to be perfectly linear for both skewness and kurtosis. This is unexpected, since it is not apparent from the expressions for the full- and cut-sky skewness and excess kurtosis

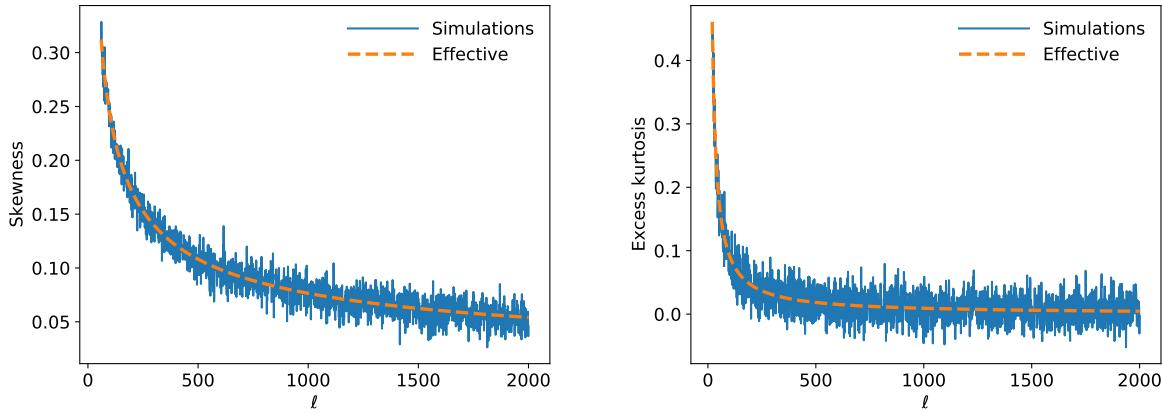


Figure 4.9. Validation of the extrapolation in Figure 4.8 against 63 100 simulated cut-sky realisations.

in Equations (4.16)–(4.17) and (4.20)–(4.21). However, additional tests have confirmed that it holds for all ten auto-spectra in this setup. It appears that the cut-sky skewness and excess kurtosis as a function of ℓ are simply linear transformations of their full-sky counterparts, with the slope of the transformation depending on the details of the mask.

In the right panel of Figure 4.8, this linear mapping is extrapolated to $\ell = 2000$. Since this is a large extrapolation, it is verified in Figure 4.9 by comparing to the sample skewness and kurtosis from 63 100 simulated cut-sky realisations of a single field. It is clearly an excellent fit.

Finally, the impact of this additional non-Gaussianity of the marginal distributions on the cut sky is tested by applying an adjusted Wishart likelihood having the correct amount of cut-sky non-Gaussianity in its marginal distributions, which is obtained by replacing ℓ in the likelihood with ℓ_{eff} . The kurtosis mapping is used here, since it gives a lower ℓ_{eff} for a given ℓ (Figure 4.8) and is therefore a more pessimistic choice. The adjusted likelihood replaces Equation (4.3) with

$$\widehat{\mathbf{C}}'_{\ell} \sim \mathcal{W} \left(\nu = 2\ell_{\text{eff}} + 1, \mathbf{V} = \frac{\mathbf{C}_{\ell}}{2\ell_{\text{eff}} + 1} \right). \quad (4.22)$$

Note that each observed ℓ still depends on the same ℓ in the theory power spectra. This means that each C_{ℓ} will retain the correct sensitivity to cosmological parameters, enabling a test of the impact of an increased amount of non-Gaussianity for the same cosmological constraining power. The marginal distributions of the auto- C_{ℓ} s in $\widehat{\mathbf{C}}'_{\ell}$ are gamma distributions, with parameters

$$\widehat{C}'_{\ell} \sim \Gamma \left(k = \frac{2\ell_{\text{eff}} + 1}{2}, \theta = \frac{2C_{\ell}}{2\ell_{\text{eff}} + 1} \right), \quad (4.23)$$

and therefore—from Equations (4.16)–(4.17)—will have the same amount of skewness and

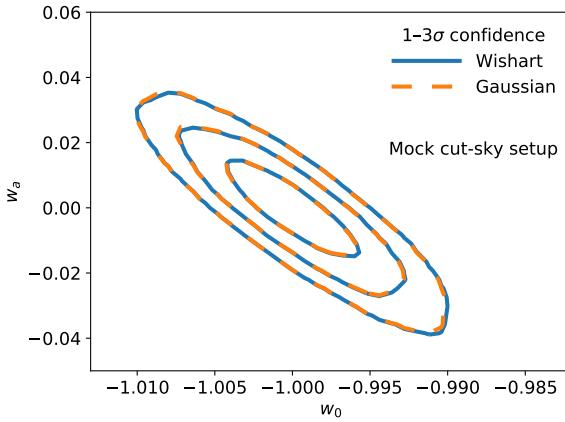


Figure 4.10. Posterior distributions obtained from a mock observation designed to have the same amount of non-Gaussianity in its marginal distributions as the cut-sky likelihood.

excess kurtosis as the cut-sky likelihood (in fact a slightly higher amount of skewness, since the kurtosis $\ell \rightarrow \ell_{\text{eff}}$ mapping is used).

The corresponding Gaussian likelihood has the same mean and covariance as the adjusted Wishart likelihood. Its mean, therefore, is unchanged from the full-sky case, while the covariance is

$$\text{Cov} \left(\hat{C}_\ell^{\alpha\beta}, \hat{C}_{\ell'}^{\gamma\epsilon} \right) = \frac{\delta_{\ell\ell'}}{2\ell_{\text{eff}} + 1} \left(C_\ell^{\alpha\gamma} C_\ell^{\beta\epsilon} + C_\ell^{\alpha\epsilon} C_\ell^{\beta\gamma} \right). \quad (4.24)$$

A mock observation following Equation (4.22) was generated by sampling directly from the Wishart distribution using the SciPy implementation of the Wishart variate generating algorithm from Smith & Hocking (1972). Using this observation, a two-parameter likelihood analysis was conducted with the adjusted Wishart and Gaussian likelihoods. The resulting posterior distribution is shown in Figure 4.10. There is very good agreement between the two likelihoods. Although small deviations are visible, this is highly likely to be an acceptable level of inaccuracy.

We can therefore conclude that the additional non-Gaussianity in the marginal distributions of the cut-sky likelihood compared to the full-sky likelihood is insufficient to introduce significant inaccuracy into the results obtained using a Gaussian likelihood.

4.3.2. Cut sky: Effect on dependence structure

To study the cut-sky dependence structure it is necessary to rely on simulations. 50 000 simulated observations were generated, following the method described in Section 4.2.2.2 with two differences: first, the observed power spectra were measured for each realisation both before and after multiplication at the map level by the *Euclid*-like mask; 10 logarithmically spaced bandpowers from $\ell = 2$ to 2000 were then formed, weighted following Equation 20 of Hivon et al. (2002).

4.3.2.1. Mutual information

Dependence between two data elements can be quantified using mutual information (MI). MI is defined as the Kullback–Leibler (KL) divergence D_{KL} of the joint distribution of two variables from the product of their marginal distributions,

$$I(X, Y) = D_{\text{KL}} \left(P_{(X,Y)} \middle\| P_X \otimes P_Y \right), \quad (4.25)$$

where the KL divergence for continuous distributions is

$$D_{\text{KL}} \left(P_{(X,Y)} \middle\| P_X \otimes P_Y \right) = \iint dx dy p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right). \quad (4.26)$$

If X and Y are independent, their joint distribution factorises and the MI vanishes. If they are not independent, they will have a positive MI. In practice, however, MI estimation from a finite number of samples may return a negative value.

To isolate non-Gaussian dependence, a whitening procedure is first applied to remove linear correlations. Linear correlations are those which are fully described by a covariance (or equivalently, correlation) matrix. Since the dependence structure in a multivariate Gaussian distribution is also fully described by its covariance matrix—such that the components of a multivariate Gaussian with diagonal covariance are independent—removing linear correlations removes all Gaussian dependence. This whitening follows the same process as Sellentin & Heavens (2018), Sellentin, Heymans & Harnois-Déraps (2018), Diaz Rivero & Dvorkin (2020) and Louca & Sellentin (2020): each pair of data elements is whitened separately using a Cholesky whitening procedure followed by a mean subtraction. The result is a whitened pair having a mean of zero and a covariance matrix of the identity matrix. Each pair is whitened separately so that pairs are still identifiable, allowing the study of the behaviour of pairs with specific relationships.

For each whitened pair, MI is estimated using the Non-parametric Entropy Estimation Toolbox (**NPEET**)² (Ver Steeg 2014). The **NPEET** MI estimator implements a k -nearest neighbours method described in Kraskov, Stögbauer & Grassberger (2004). The default parameters of $k = 3$ and log base 2 are used in Equation (4.26).

Figure 4.11 shows the distribution of pairwise whitened MI compared between full-sky and cut-sky bandpowers. Most of the pairs of elements are found in the part of the distribution centred around zero, indicating no detected non-Gaussian dependence. This is more clearly seen in Figure 4.12, in which each of the full-sky and cut-sky samples is compared to an equivalent sample drawn from a multivariate Gaussian distribution having the same mean and covariance. Non-Gaussian dependence is exhibited by the pairs of elements found in the tail, in this case with $\text{MI} \gtrsim 0.02$. This tail contains only a small fraction of pairs in both

² <https://github.com/gregversteeg/NPEET>

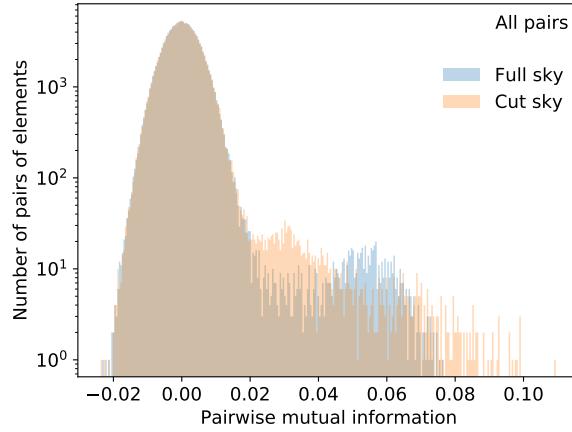


Figure 4.11. Comparison of non-Gaussian dependence between full-sky and cut-sky likelihoods. Non-Gaussian dependence is quantified by pairwise mutual information after whitening. The mass centred around zero on the x -axis represents Gaussian dependence, with the tail (here $\gtrsim 0.02$) representing non-Gaussian dependence.

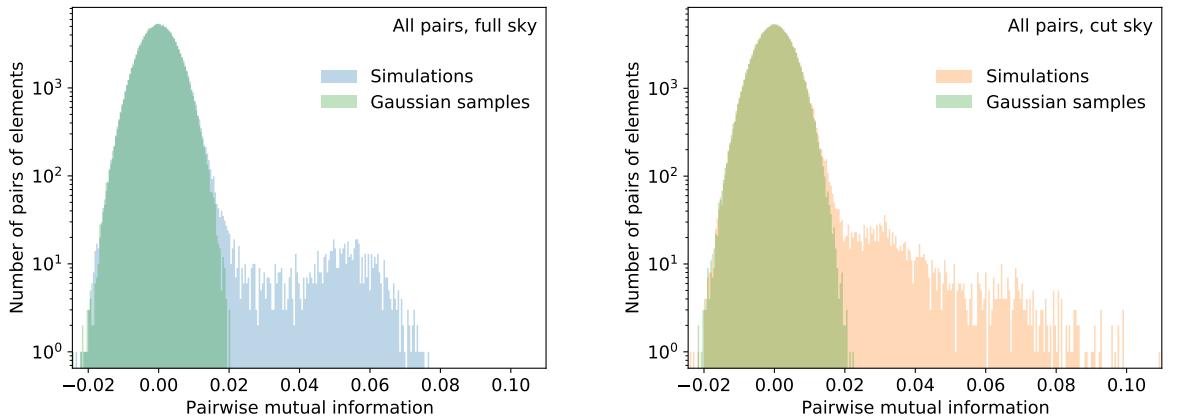


Figure 4.12. Comparison of non-Gaussian dependence between full-sky (left) and cut-sky (right) likelihoods. Non-Gaussian dependence is quantified by pairwise mutual information after whitening. In both panels the corresponding distribution for pure Gaussian dependence is shown, which is centred around zero on the x -axis. The tail (here $\gtrsim 0.02$) represents non-Gaussian dependence.

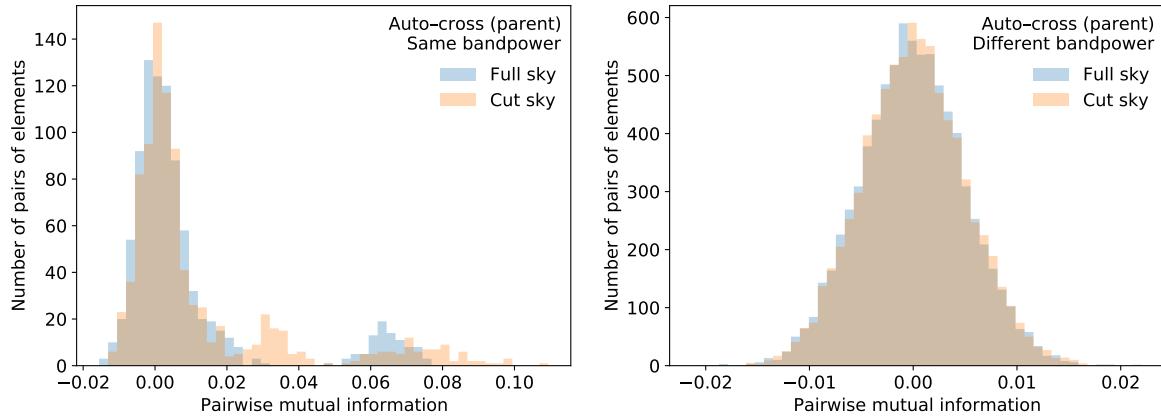


Figure 4.13. Comparison of non-Gaussian dependence between full-sky and cut-sky likelihoods, for two specific populations of data element pairs. The left panel shows pairs containing the equivalent bandpower of one cross-spectrum and one of its ‘parent’ auto-spectra (e.g. P_b^{12} and P_b^{11}). The right panel is for different-bandpower pairs in the same combinations of spectra (e.g. P_b^{12} and $P_{b'}^{11}$). Non-Gaussian dependence is quantified by pairwise mutual information after whitening. The bulk centred around zero on the x -axis represents Gaussian dependence, with the tail representing non-Gaussian dependence. The absence of this tail in the right panel indicates that those pairs exhibit only Gaussian dependence.

cases, but with a slight excess for the cut-sky sample: 0.93 per cent of cut-sky pairs have $\text{MI} > 0.02$, compared to 0.65 per cent of full-sky pairs. This is also evident in the small visible excess of cut-sky pairs in Figure 4.11.

To investigate the origin of this small excess in non-Gaussian dependence for the cut-sky sample relative to the full-sky sample, each sample is split into different pair populations, corresponding to particular relationships between data elements. Non-Gaussian dependence is almost exclusively found in pairs containing the same bandpower across correlated fields. The strongest such case is shown in the left panel of Figure 4.13, which shows pairs containing one bandpower from a cross-spectrum and the same bandpower from one of its ‘parent’ auto-spectra, i.e. the auto-spectrum of one of the two fields between which the cross-spectrum is describing the correlation. While most pairs still appear consistent with zero, there is a significant tail of non-Gaussian dependence, which is slightly larger for the cut-sky sample. This tail is not found when looking at pairs of different bandpowers between the same two spectra, shown in the right panel of Figure 4.13. A similar behaviour is found in other same-bandpower pairs, which is strongest when the two spectra in the pair relate directly to the same underlying field; for example, two ‘sibling’ cross-spectra which share one parent auto-spectrum. In most such pair populations, there is a slight excess of non-Gaussian dependence for the cut-sky sample.

In contrast, it turns out that the dependence between bandpowers in the same spectrum known to be induced by a cut sky in fact comprises almost purely linear correlations. This is shown in Figure 4.14, which compares these pairs before and after the whitening process.

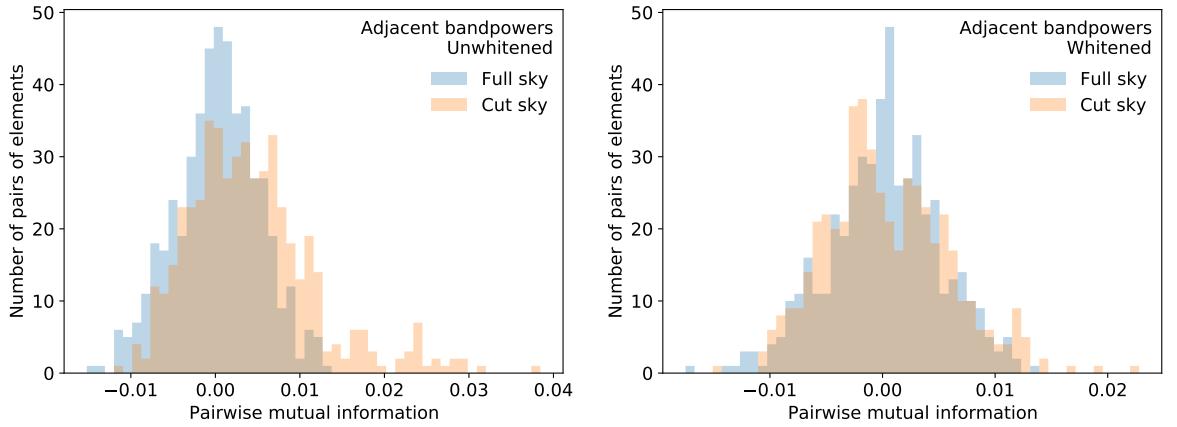


Figure 4.14. Full-sky and cut-sky distributions of pairwise mutual information (MI) before (left) and after (right) pairwise whitening, for pairs of adjacent bandpowers in the same spectrum (e.g. P_b^{12} and P_{b+1}^{12}). Prior to whitening, MI captures all dependence; in this case there is a substantial excess of dependence in the cut-sky likelihood. Whitening removes Gaussian dependence such that after whitening, MI only captures non-Gaussian dependence; in this case there is little or no excess dependence in the cut-sky likelihood. This demonstrates that the additional dependence between adjacent bandpowers induced by a mask is mostly or wholly Gaussian.

The left panel shows the unwhitened result, which includes linear correlations, showing an expected cut-sky excess. After whitening, shown in the right panel, this excess is almost entirely removed.

4.3.2.2. Impact of additional non-Gaussian dependence

The above section has shown that there is a small excess in non-Gaussian dependence in the cut-sky likelihood compared to the full-sky likelihood. As was done for the marginal distributions in Section 4.3.1.1, the potential impact of this additional non-Gaussianity on the accuracy of constraints obtained using the Gaussian likelihood is now investigated.

On closer inspection it turns out that the increased non-Gaussian dependence in the cut-sky likelihood is in fact an ℓ -dependent effect. This is demonstrated in Figure 4.15, which shows whitened MI as a function of ℓ for same-bandpower pairs, which as discussed above are those which exhibit non-Gaussian dependence. Non-zero MI appears to be restricted only to the lowest bandpowers, with a small excess for the cut sky. This ℓ dependence resembles that of the skewness and kurtosis of the marginal distributions, and implies that the mock cut-sky data vector and likelihood that was developed in Section 4.3.1.1 using the $\ell \rightarrow \ell_{\text{eff}}$ mapping process should have higher MI—indicating more non-Gaussian dependence—than the full-sky data and likelihood tested in Section 4.2. The average MI in the mock cut-sky setup can be conservatively estimated by taking the full-sky MI sample and replacing the MI value of each same-bandpower pair at any ℓ with that of its corresponding ℓ_{eff} , interpolating the full-sky MI-vs.- ℓ curve shown in Figure 4.15. The MI

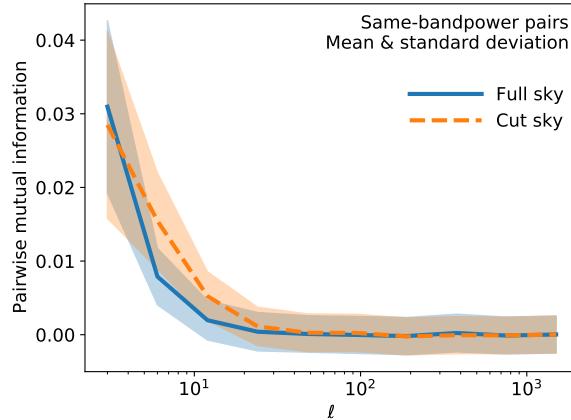


Figure 4.15. Non-Gaussian dependence as a function of ℓ , compared between full-sky and cut-sky likelihoods. Non-Gaussian dependence is quantified by pairwise mutual information after whitening, and is here averaged over all same-bandpower pairs in a given ℓ bin, with the shaded region containing one standard deviation.

of different-bandpower pairs is left unchanged. This gives an average MI of 9.0×10^{-4} , which compares to 4.1×10^{-4} for the full-sky sample and 5.2×10^{-4} for the cut-sky sample. So the mock cut-sky sample and likelihood has roughly 80 per cent more non-Gaussian dependence than the true cut-sky sample and likelihood, and yet the resulting posterior distribution from the Gaussian likelihood in Figure 4.10 is still extremely accurate. Therefore, we can conclude that the impact of additional non-Gaussian dependence on the cut sky is negligible.

4.3.2.3. Transcovariance

Transcovariance is a measure of non-Gaussianity of a distribution introduced in Sellentin & Heavens (2018) and subsequently used in Sellentin, Heymans & Harnois-Déraps (2018), Louca & Sellentin (2020) and Diaz Rivero & Dvorkin (2020). This section will follow the latter three papers in considering only the additive transcovariance S^+ , which is defined as

$$S^+ = \frac{1}{B} \sum_{b=1}^B [\mathcal{H}_b - \mathcal{N}_b(0, 2)]^2, \quad (4.27)$$

where the sum is over the bins b of a histogram \mathcal{H} of the sum of two data elements after whitening, and $\mathcal{N}(0, 2)$ is the expected histogram of a univariate Gaussian distribution with mean 0 and variance 2.

S^+ is a measure of non-Gaussianity, because if the two data elements were bivariate Gaussian distributed, then their sum after whitening would be univariate Gaussian with mean 0 and variance 2, and so the expectation $E[\mathcal{H}_b - \mathcal{N}_b(0, 2)]$ would vanish. S^+ has sometimes been described as a measure of “non-Gaussian correlations”, but in practice it is sensitive to non-Gaussianity of both the marginals and the dependence. For example, if

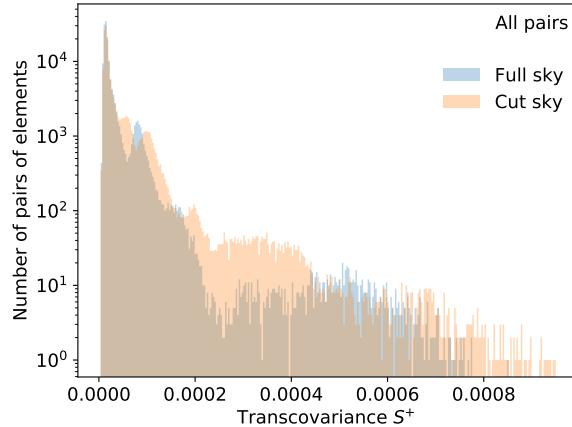


Figure 4.16. Full-sky and cut-sky distributions of pairwise additive transcovariance after pairwise whitening. Transcovariance is an alternative measure of non-Gaussianity, discussed in Section 4.3.2.3.

two data elements each had a non-Gaussian marginal distribution but their dependence was purely linear correlation (or more simply, if they were independent), then their dependence would vanish with the whitening procedure and yet they would return a non-zero S^+ value because their sum would not, in general, follow a Gaussian distribution. As such, S^+ is a holistic measure of non-Gaussianity, which is a useful property in many applications. However, it is for this reason that it is not used here as the main test of non-Gaussian dependence, as it would not allow the separate consideration of the marginal distributions and dependence structure of the likelihood.

For completeness, the distributions of additive transcovariance for the full-sky and cut-sky samples are shown in Figure 4.16. There is a much larger excess of transcovariance in the cut-sky likelihood compared to the full-sky likelihood than is seen for the MI in Figure 4.11. The fact that the transcovariance mixes the effects of non-Gaussian marginals and non-Gaussian dependence would prevent the identification of whether this is due to the marginals, the dependence or both. As an additional check, a probability integral transform was applied to each pair such that the marginal distributions were Gaussian distributed, without affecting the dependence structure, and the resulting distribution had a much smaller cut-sky excess, similar to the MI shown previously in Figure 4.11.

4.4. Non-Gaussian fields

It has been demonstrated in Section 4.2 that a Gaussian likelihood is sufficient to obtain accurate parameter constraints in a combined weak lensing and galaxy clustering analysis on the full sky, and in Section 4.3 that the additional non-Gaussianity of the cut-sky likelihood is insufficient to introduce significant inaccuracy, both provided that the observable fields may be described using Gaussian statistics. There are reasons to believe this to be a

good approximation for the purposes of this study.

First, the matter distribution is most Gaussian on linear scales, corresponding to low ℓ , and most non-Gaussian at high ℓ . But the inverse is true for the power spectrum likelihood: it is most Gaussian at high ℓ , and most non-Gaussian at low ℓ . While this behaviour has been demonstrated in the previous sections for Gaussian fields, it can be expected to hold generally, as the number of $a_{\ell m}$ s contributing to each C_ℓ estimate increases with ℓ regardless of the statistics of the field. The largest contribution to potential inaccuracy in the Gaussian likelihood therefore comes from linear scales, where the observable fields are well described as Gaussian. Additionally, the presence of shape noise causes two further effects: it decreases the non-Gaussianity of the fields on all scales, and relatively upweights the contribution of large scales to the overall constraining power, as discussed in Section 4.2.4.2. Both of these effects will increase the accuracy of the Gaussian fields assumption. Finally, the process of going from galaxy catalogues to power estimates involves first averaging over galaxies in each pixel, followed by a spherical harmonic transform, both of which may be expected to approximately Gaussianise the $a_{\ell m}$ s following the central limit theorem. The latter principle has been tested in simple simulations using HEALPix spherical harmonic transforms of arbitrary non-Gaussian fields and found to hold. Gaussian $a_{\ell m}$ s in turn imply approximately gamma-distributed auto- C_ℓ estimates. However, the degrees of freedom in these gamma distributions may be reduced (and hence the non-Gaussianity increased) if the $a_{\ell m}$ s are correlated, similar to what happens on the cut sky. This idea was tested in Taylor et al. (2019), which found no detectable difference between C_ℓ distributions measured from Gaussian and lognormal simulations, which have been shown to well approximate real weak lensing data (Taruya et al. 2002; Hilbert, Hartlap & Schneider 2011; Clerkin et al. 2016).

However, non-Gaussian fields will introduce additional contributions to the C_ℓ covariance, which have been neglected here. Specifically, there is a contribution arising from four-point correlations within a survey volume—often referred to as the connected non-Gaussian covariance—and a generally larger contribution arising from the dependence of such correlations on unmeasured super-survey modes—commonly termed super-sample covariance. This topic is explored in detail in Chapter 5. (See also e.g. Scoccimarro, Zaldarriaga & Hui 1999; Cooray & Hu 2001; Takada & Bridle 2007; Takada & Hu 2013; Li, Hu & Takada 2014; Barreira, Krause & Schmidt 2018a,b.) These additional covariance contributions predominantly affect high ℓ , which will have the effect of relatively upweighting the low- ℓ regime where the Gaussian likelihood is least accurate. However, Section 4.2.4.2 showed that the Gaussian likelihood still performs well even at extremely low ℓ , so we should not expect this to outweigh the other factors outlined above, and we can expect the overall impact of non-Gaussian fields on the accuracy of a Gaussian likelihood to be negligible. Nevertheless, here the same techniques as in Section 4.3 are applied to test the non-Gaussianity detected in a more realistic set of weak lensing simulations.

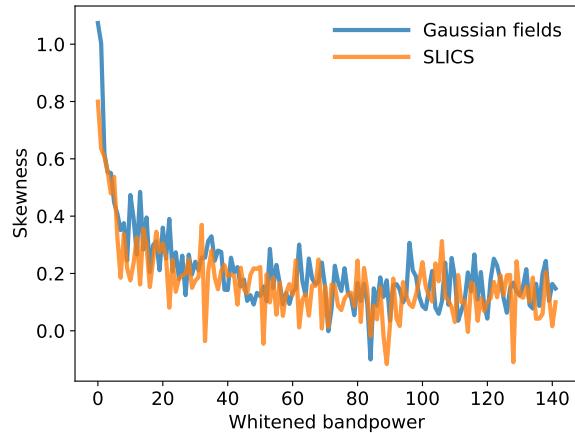


Figure 4.17. Skewness of SLICS bandpowers compared to the Gaussian fields sample, after whitening to remove the effect of linear correlations.

4.4.1. Non-Gaussian fields: Simulations

The tests presented in this section use the SLICS³, which are independent N-body weak lensing simulations, described in detail in Harnois-Déraps & Van Waerbeke (2015) and Harnois-Déraps et al. (2018). The only available tomographic power spectra were the weak lensing convergence power spectra from the KiDS-450-like setup used in Hildebrandt et al. (2017). These are flat-sky linearly-spaced bandpowers for auto-spectra only, produced from 948 independent realisations of 60 deg^2 sky patches. Here a value of $\ell_{\max} = 5000$ is used, in order to include non-linear scales.

An equivalent batch of Gaussian-field simulations were generated using `pymaster`, the Python implementation of `NaMaster`⁴ (Alonso, Sanchez & Slosar 2019), using a KiDS-450-like setup with four tomographic bins following the specification in Table 1 of Hildebrandt et al. (2017).

Following the process in Section 4.3, the non-Gaussianity of the marginals and dependence in the distributions from the SLICS is now compared to the Gaussian field sample.

4.4.2. Non-Gaussian fields: Effect on marginal distributions

Figure 4.17 shows the skewness of the SLICS compared to the Gaussian field sample, averaged over the four redshift bins. Each sample was whitened prior to calculating skewness, because significant linear correlations were found to be present in the SLICS data that were not present in the Gaussian field simulations. These correlations are likely to be real rather than an artefact, since the SLICS were designed and validated specifically for covariance estimation (Harnois-Déraps & Van Waerbeke 2015). However, what matters for the accu-

³ <https://slics.roe.ac.uk>

⁴ <https://github.com/LSSTDESC/NaMaster>

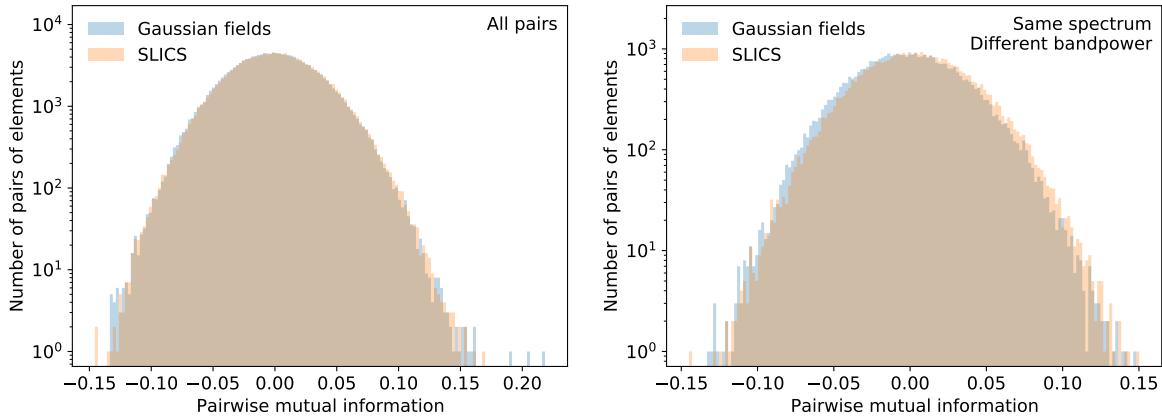


Figure 4.18. Comparison of non-Gaussian dependence between the SLICS weak lensing simulations and a similar sample of Gaussian fields. Non-Gaussian dependence is quantified by pairwise mutual information after whitening. The left panel shows all pairs of data elements, while the right panel shows only those containing different bandpower pairs in the same spectrum. This population exhibits a possible small excess in non-Gaussian dependence.

racy of a multivariate Gaussian distribution is the non-Gaussianity of the marginals after whitening, since the Gaussian PDF effectively whitens the data vector itself. As shown in Figure 4.17, the skewness after whitening is consistent to within the level of the noise, with a possible slight excess for the Gaussian fields. Similar consistency is found for the excess kurtosis. We can therefore conclude that there is no evidence of additional non-Gaussianity of the marginals for realistic non-Gaussian weak lensing fields.

4.4.3. Non-Gaussian fields: Effect on dependence structure

Following the procedure described in Section 4.3.2, pairwise mutual information (MI) is measured after a pairwise whitening process, and this is compared between the SLICS and the Gaussian field sample. The overall MI distributions, shown in the left panel of Figure 4.18, are almost indistinguishable, but there is a very small excess for the SLICS. By splitting pairs of data elements into populations depending on their relationship, it is revealed that this excess is due to a particular population: different bandpowers in the same spectrum, shown in the right panel of Figure 4.18. There is no apparent redshift dependence in this behaviour, nor does it have any apparent ℓ -dependent structure: Figure 4.19 shows the matrix of pairwise MI compared between the two samples. Both samples appear consistent with noise, with a slightly higher noise level in the SLICS. Whether this is a real or spurious effect is unknown; however, we can expect its effect on the accuracy of the Gaussian likelihood to be negligible: the average MI for the SLICS is 1.9×10^{-4} , far below the 9.0×10^{-4} that was found to cause negligible inaccuracy in constraints obtained using the Gaussian likelihood in Figure 4.10. Therefore, no evidence is found to suggest the conclusions drawn from the Gaussian field tests in Section 4.2 and Section 4.3 should not hold

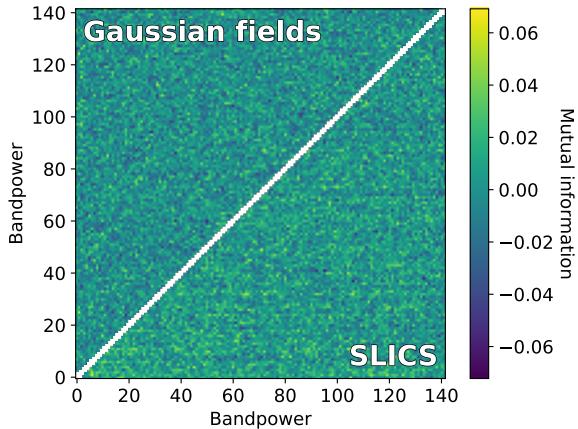


Figure 4.19. Matrix of pairwise mutual information after pairwise whitening, averaged over four redshift bins, compared between the SLICS and the Gaussian fields sample.

for real weak lensing fields.

4.5. Conclusions

It is well established that the true likelihood of weak lensing two-point statistics is non-Gaussian (Sellentin & Heavens 2018; Sellentin, Heymans & Harnois-Déraps 2018; Diaz Rivero & Dvorkin 2020; Louca & Sellentin 2020), and yet contemporary analyses routinely neglect this and assume a Gaussian likelihood (Troxel et al. 2018; Hikage et al. 2019; Joachimi et al. 2021). The work presented in this chapter has tested the impact of assuming a Gaussian likelihood for a *Euclid*-like combined power spectrum analysis of weak lensing, galaxy clustering and their cross-correlation, on the inferred posterior distributions of dark energy parameters.

In Section 4.2 it was found that on the full sky, the Gaussian likelihood returns the correct posterior maximum, two-dimensional contours and one-dimensional posterior probability density. This holds both when all other parameters are fixed or when marginalising over a third parameter, and for any choice of fiducial cosmology consistent with the data. The Gaussian likelihood is even a good approximation at low ℓ , where the true likelihood is most non-Gaussian.

It was shown in Section 4.3 that a sky cut increases the non-Gaussianity of both the marginal distributions and dependence structure of the likelihood. However, by generating a mock cut-sky data vector and likelihood with the appropriate amount of non-Gaussianity in both cases, it was shown that this additional non-Gaussianity introduces only negligible additional inaccuracy into the posterior parameter constraints obtained using the Gaussian likelihood.

The results presented in Section 4.2 and Section 4.3 were obtained under the assumption

of Gaussian fields. It was argued in Section 4.4 that this is a sufficient approximation for the purposes of this analysis. Nevertheless, results obtained under this approximation were compared to those obtained using an equivalent set of N-body weak lensing simulations, and no evidence was found of significant additional non-Gaussianity of the power spectrum likelihood.

These results indicate that a Gaussian likelihood will be sufficient for robust cosmological inference with power spectra from stage IV weak lensing surveys such as *Euclid*. This conclusion is further supported by the results obtained in Taylor et al. (2019), which found no significant difference in parameter constraints obtained using a Gaussian likelihood compared to a likelihood-free approach. We cannot automatically extend this conclusion to the correlation function, which has a more complicated behaviour due to the mixing of scales (Sellentin & Heavens 2018). Lin et al. (2020) have found that a Gaussian likelihood is likely to be sufficiently accurate for parameter inference from LSST data. However, the disagreement between that result and that of Hartlap et al. (2009), who found that the assumption of a Gaussian correlation function likelihood introduced significant inaccuracy in parameter constraints from a weak lensing analysis of the Chandra Deep Field South, remains to be fully understood.

Chapter 5

Covariance of weak lensing pseudo- C_ℓ estimates

5.1. Introduction

There are currently many unanswered questions in cosmology, including the origin of the accelerating expansion of the Universe and apparent tensions within the dominant Λ CDM model. As described in Chapter 1, one of the most promising tools with which to make progress on these questions in the coming years is the analysis of weak gravitational lensing of distant galaxies by large-scale structure, also known as cosmic shear. The upcoming ESA *Euclid* space mission, as well as other surveys such as those with the Vera C. Rubin Observatory in Chile and the Square Kilometre Array radio observatory in Australia and South Africa, will observe over a billion galaxies, which is expected to lead to unprecedented precision on cosmological constraints—a more than an order of magnitude increase over the previous generation of experiments (Harrison et al. 2016). In order to obtain reliable results, this precision is necessarily accompanied by a requirement to understand all elements of an analysis pipeline to an equally unprecedented degree, including the interplay between the likelihood and estimator effects.

Continuing from Chapters 3 and 4, this chapter is concerned specifically with pseudo- C_ℓ estimators, which were introduced in Chapter 2. Pseudo- C_ℓ estimators have been used previously for the analysis of weak lensing data from the Hyper-Suprime Cam Subaru Strategic Program in Hikage et al. (2019) and the Dark Energy Survey (DES) in Camacho et al. (2021) and will be used in the analysis of future *Euclid* data (Loureiro et al. 2021). It was shown in Chapter 4 that a Gaussian likelihood is sufficient to obtain accurate cosmological results from weak lensing pseudo- C_ℓ estimates. An important ingredient for a Gaussian likelihood is the covariance matrix, so this chapter focuses on the calculation of a cosmic shear pseudo- C_ℓ covariance matrix.

The problem of calculating covariance matrices for weak lensing has been extensively discussed in the literature, ranging from analytic or semi-analytic approaches (Cooray & Hu 2001; Schneider et al. 2002; Joachimi, Schneider & Eifler 2008b; Takada & Jain 2009; Pielorz et al. 2010; Hilbert, Hartlap & Schneider 2011; Barreira, Krause & Schmidt 2018a; Hall & Taylor 2019; Gouyou Beauchamps et al. 2022) through to estimation from simulations (Sato et al. 2011; Harnois-Déraps & Van Waerbeke 2015; Sellentin & Heavens 2016b,a; Harnois-Déraps et al. 2018; Harnois-Déraps, Giblin & Joachimi 2019; Sgier et al. 2019; Schneider et

al. 2020). This chapter extends this work to focus specifically on the covariance of pseudo- C_ℓ estimates, for which coupling between modes occurs due to the effect of incomplete sky coverage. This effect is in addition to the non-Gaussian mode coupling that is inherent in weak lensing data as a result of non-linear structure growth, and which is known to be important for parameter inference (Sato & Nishimichi 2013; Barreira, Krause & Schmidt 2018b).

In Section 5.2 the different Gaussian and non-Gaussian components of the cosmic shear pseudo- C_ℓ covariance and their implementation in existing publicly available code are described. This theoretical covariance is compared to that measured from publicly available weak lensing simulations in Section 5.3. Section 5.4 examines the relative importance of the different covariance contributions and how this depends on the mask, which describes the details of sky coverage. This part of the analysis shares some similarities with that of Barreira, Krause & Schmidt (2018b), who also studied the relative importance of the different contributions to the cosmic shear covariance for a *Euclid*-like survey and concluded that the ‘connected non-Gaussian’ component (see Section 5.2) can be neglected for only a $\lesssim 5$ per cent underestimation in single-parameter 1σ errors. However, this chapter is specifically focused on pseudo- C_ℓ estimates, for which the survey mask mixes power between all multipoles and induces correlations even for Gaussian fields, which for many covariance elements dominate over other sources of correlation (see Section 5.3). This effect was not included in the analysis of Barreira, Krause & Schmidt (2018b), who assumed a diagonal Gaussian covariance, and its inclusion may lead to different conclusions about the relative importance of the different contributions to the covariance. The conclusions of this work are discussed in Section 5.5.

5.2. Cosmic shear power spectrum covariance contributions

There are three contributions to the cosmic shear power spectrum covariance, which are summarised below. A thorough theoretical background and derivation may be found in Barreira, Krause & Schmidt (2018a) and the other references provided both therein and below.

Starting in three-dimensional space, the covariance of the matter power spectrum receives two contributions: one that depends on the matter power spectrum itself, and one that depends on a particular (‘parallelogram’) configuration of the matter trispectrum that corresponds to the Fourier transform of the connected four-point correlation function. For a Gaussian matter distribution, only the first contribution is non-vanishing, and hence it is commonly referred to as the ‘Gaussian covariance’, which will be used in this chapter. (It is also sometimes referred to as the ‘disconnected’ covariance.) Following Barreira, Krause & Schmidt (2018a), the second contribution is referred to as the ‘connected non-Gaussian’ component.

However, for any realistic finite-volume survey such as *Euclid*, the observed matter power spectrum is convolved with a three-dimensional window function. While the Gaussian and non-Gaussian terms remain distinct, this has the effect of introducing additional non-Gaussian coupling between large-scale modes outside the survey and small-scale modes within the survey. This is commonly known as ‘super-sample’ (originally ‘beat-coupling’; Hamilton, Rimes & Scoccimarro 2006) covariance, and physically can be explained by the fact that unobservable large-scale modes within which the survey is embedded can influence the rate of small-scale non-linear structure growth, and therefore also the strength of coupling between small-scale modes (Takada & Hu 2013; Barreira, Krause & Schmidt 2018a). Perhaps counter-intuitively, it turns out that this is generally the dominant source of non-Gaussian covariance (Hamilton, Rimes & Scoccimarro 2006; Barreira, Krause & Schmidt 2018b).

Progressing to projected two-point statistics such as cosmic shear angular power spectra, the same three components—Gaussian, super-sample and connected non-Gaussian—contribute to the covariance. Strictly speaking, the separation of the super-sample and connected non-Gaussian components is only exact under the Limber approximation (Barreira, Krause & Schmidt 2018a), but the inaccuracy of the Limber approximation is only relevant on very large scales (very low multipoles, $\ell \lesssim 20$) where non-Gaussian correlations are small.

The calculation of the three cosmic shear covariance components are each now discussed in turn.

5.2.1. Gaussian covariance

To calculate the Gaussian covariance, the ‘improved narrow kernel approximation’ method (Nicola et al. 2021) was used, which was implemented using the publicly available code `NaMaster` (Alonso, Sanchez & Slosar 2019; García-García, Alonso & Bellini 2019). Further details and some background on this method are provided as follows.

The Gaussian covariance component of a general statistically isotropic field on the sphere is equivalent to the total covariance of a Gaussian field with the same power spectrum. The analytic covariance of pseudo- C_ℓ estimates on Gaussian fields has been well studied in the cosmic microwave background literature (Efstathiou 2004; Challinor & Chon 2005; Brown, Castro & Taylor 2005) as well as in the context of weak lensing (García-García, Alonso & Bellini 2019; Nicola et al. 2021). The exact Gaussian pseudo- C_ℓ covariance can be written down analytically, and includes terms of the following form (e.g. Brown, Castro & Taylor 2005):

$$\begin{aligned} \text{Cov}\left(\tilde{C}_\ell, \tilde{C}_{\ell'}\right) = & \sum_{\substack{m, m' \\ \ell_1, \ell_2 \\ m_1, m_2}} W_{\ell\ell_1}^{mm_1} \left(W_{\ell'\ell_1}^{m'm_1}\right)^* W_{\ell'\ell_2}^{m'm_2} \left(W_{\ell\ell_2}^{mm_2}\right)^* C_{\ell_1} C_{\ell_2} \\ & + \text{similar terms}, \end{aligned} \quad (5.1)$$

where the harmonic space window functions W are given in Equation (8) of Brown, Castro & Taylor (2005), and the ‘similar terms’ involve different combinations of power spectra depending on the situation and spins being considered (Hansen & Górski 2003; Challinor & Chon 2005).

The evaluation of Equation (5.1) requires $\mathcal{O}(\ell_{\max}^6)$ operations per term, so it is impractical to evaluate exactly and in practice approximations are used. These commonly involve substitutions of the following kind (Efstathiou 2004; Brown, Castro & Taylor 2005; García-García, Alonso & Bellini 2019):

$$C_{\ell_1} C_{\ell_2} \rightarrow C_\ell C_{\ell'}, \quad (5.2)$$

which allows the power spectrum dependence to be brought out of the sums in Equation (5.1). This means that the coefficients in the similar terms are now all the same (except for any possible spin dependence, or if different fields use different masks). Symmetry properties of the harmonic space window function allow the calculation of these coefficients to be further simplified, to the point where the covariance can be evaluated in a reasonable time. In essence, the approximation in Equation (5.2) assumes that the power spectrum is constant over the region around a given ℓ in which the window function is non-negligible. This will be accurate as long as the window function is sufficiently sharply peaked, and therefore this approximation is often known as the ‘narrow kernel approximation’ (NKA). A generalised version of the NKA is described in García-García, Alonso & Bellini (2019) and implemented in NaMaster, which supports an arbitrary number of correlated spin-0 and spin-2 fields and has both curved-sky and flat-sky support. For this work the curved-sky spin-2 version was used, which naturally accounts for E - B leakage (here assuming noise-only B -modes). By default, NaMaster provides the covariance of deconvolved pseudo- C_ℓ estimates, but here the `coupled=True` option is set to instead obtain the covariance of ‘raw’, un-deconvolved estimates such as those produced by the pseudo- C_ℓ estimator developed for *Euclid*, which is described in Loureiro et al. (2021). No E - B purification or noise debiasing was applied.

Nicola et al. (2021) introduced a small modification to the NKA that turns out to significantly increase its accuracy, which they refer to as the improved NKA. It involves simply replacing each C_ℓ in the standard NKA by its mode-coupled counterpart,

$$C_\ell \rightarrow \frac{\langle \tilde{C}_\ell \rangle}{f_{\text{sky}}} = \frac{\sum_{\ell'} \mathbf{M}_{\ell\ell'} C_{\ell'}}{f_{\text{sky}}}, \quad (5.3)$$

where \mathbf{M} is the usual pseudo- C_ℓ mixing or mode-coupling matrix (see Brown, Castro & Taylor 2005 for a full derivation of its calculation), and the division by the sky fraction f_{sky}

is to avoid double-counting the loss of power on the cut sky. NaMaster includes the functionality to calculate the mixing matrix and to apply it to power spectra to calculate $\langle \tilde{C}_\ell \rangle$, so the extension to the improved NKA is trivial. The recent pseudo- C_ℓ analysis of DES Year 1 observations in Camacho et al. (2021) provided the first application of the improved NKA to real data, but takes the approach of deconvolution and noise subtraction to obtain unbiased estimates of the underlying power spectrum, unlike the forward-modelling approach taken in this chapter.

5.2.2. Super-sample covariance

To calculate the super-sample covariance contribution, the publicly available code `CosmoLike`¹ (Krause & Eifler 2017) was used; specifically, an adapted version of the `CosmoCov`² correlation function covariance package (Fang, Eifler & Krause 2020), which obtains the real-space non-Gaussian covariance as a transform of the harmonic space covariance. `CosmoCov` has been used for the DES Year 1 and Year 3 cosmological analyses (Krause et al. 2017, 2021; Friedrich et al. 2021), as well as for the non-Gaussian covariance in the Year 1 pseudo- C_ℓ analysis in Camacho et al. (2021). The code adapted for this work to expose the harmonic space covariance directly is available online.³

`CosmoLike` calculates the super-sample covariance using the approach introduced in Takada & Hu (2013), by considering the response of the small-scale non-linear matter power spectrum to changes in the mean linear density field (see also Chiang et al. 2014; Li, Hu & Takada 2014). The response of the non-linear matter power spectrum is evaluated using a halo model, with the details given in Equations (A7)–(A11) of Krause & Eifler (2017). Since that paper was published, the calculation of the survey variance $\sigma_b(z)$ in their Equation (A8) has been replaced within `CosmoLike` with the following calculation:

$$\sigma_b(z) = \frac{1}{4\pi r^2} \frac{1}{C_{\ell=0}^{\text{mask}}} \sum_{\ell=0}^{1000} (2\ell+1) C_\ell^{\text{mask}} P_{\text{lin}}\left(\frac{\ell+1/2}{r}, z\right), \quad (5.4)$$

where $r = f_\kappa(\chi(z))$, and C_ℓ^{mask} is the power spectrum of the mask. This treatment of the cut sky using the mask power spectrum was derived in Barreira, Krause & Schmidt (2018a). The value of $\ell_{\text{max}} = 1000$ in Equation (5.4) is arbitrary, but has a negligible impact in practice because the power spectrum of both masks is negligible above $\ell \sim 100$. `CosmoLike` also provides an alternative implementation of both the super-sample and connected non-Gaussian covariance using the ‘response approach’, which additionally accounts for a tidal contribution to the super-sample covariance and has been found to be more accurate than the standard implementation (Wagner et al. 2015; Barreira & Schmidt 2017a,b; Barreira, Krause & Schmidt 2018a; Schmidt et al. 2018). Here the standard `CosmoLike` implementa-

¹ <https://github.com/CosmoLike>

² <https://github.com/CosmoLike/CosmoCov>

³ https://github.com/robinupham/CosmoCov_ClCov

tion is used, as has been used by the DES Collaboration (Krause et al. 2017, 2021; Friedrich et al. 2021; Camacho et al. 2021).

As stated previously, the approach taken in this work is to forward-model the effect of the mask to obtain the covariance of raw un-deconvolved pseudo- C_ℓ estimates following Loureiro et al. (2021). The non-Gaussian covariance output from `CosmoLike` corresponds to the covariance of unbiased estimates of the underlying power spectrum \hat{C}_ℓ , and does not account for any estimator effects such as cut-sky mode coupling. Within the context of the pseudo- C_ℓ method, the closest way to interpret the covariance output from `CosmoLike` is as the covariance of deconvolved pseudo- C_ℓ estimates. This is how it is interpreted in Camacho et al. (2021), and how it is interpreted in this work. In this case, the unbiased estimates \hat{C}_ℓ may in principle be obtained from raw pseudo- C_ℓ estimates \tilde{C}_ℓ as

$$\hat{C}_\ell = \sum_{\ell'} \mathbf{M}_{\ell\ell'}^{-1} \tilde{C}_{\ell'}. \quad (5.5)$$

It follows that the covariance matrices of \hat{C}_ℓ and \tilde{C}_ℓ , denoted respectively here as $\hat{\Sigma}$ and $\tilde{\Sigma}$, are related as

$$\hat{\Sigma} = (\mathbf{M}^{-1}) \tilde{\Sigma} (\mathbf{M}^{-1})^\top. \quad (5.6)$$

The covariance as output from `CosmoLike` is interpreted as $\hat{\Sigma}$. This choice is necessarily an approximation since `CosmoLike` does not account for any estimator effects, including pseudo- C_ℓ mode coupling, but it is a necessary choice and is equivalent to that made in the DES Y1 pseudo- C_ℓ analysis in Camacho et al. (2021). In both that paper and this work (Section 5.3), the resulting covariance is compared to simulations, with good agreement. A full general non-Gaussian pseudo- C_ℓ covariance is presented in Shirasaki, Hamana & Yoshida (2015), but in practice an approximation such as the one made here is necessary. Equation (5.6) is therefore inverted to obtain the relation that was used to transform $\hat{\Sigma}$ to the raw pseudo- C_ℓ covariance, $\tilde{\Sigma}$:

$$\tilde{\Sigma} = \mathbf{M} \hat{\Sigma} \mathbf{M}^\top. \quad (5.7)$$

The spin-2 mixing matrix \mathbf{M} for this calculation was obtained using `NaMaster`.

5.2.3. Connected non-Gaussian covariance

The connected non-Gaussian covariance component was also calculated using `CosmoLike` (referred to in Krause & Eifler 2017 as the ‘non-Gaussian covariance in the absence of survey window effects’), using the same adapted version of the `CosmoCov` package that is made available at the URL provided in Section 5.2.2.

`CosmoLike` calculates the connected non-Gaussian covariance as the projected matter trispectrum. The trispectrum is calculated using a halo model, with the details given in Equations (A3)–(A6) of Krause & Eifler (2017) (see also Cooray & Sheth 2002; Takada &

Jain 2009). This is found to be suitably accurate in Section 5.3.2, and it is also shown in Section 5.4 that the contribution from the connected non-Gaussian component to the total parameter posterior error is no more than 10–20 per cent. As with the super-sample covariance component, the connected non-Gaussian covariance matrix was multiplied by the mixing matrix using Equation (5.7).

The calculation of the connected non-Gaussian covariance component is much slower than the other two components. As a result, in Section 5.4.2 an approximation is used to directly estimate the connected non-Gaussian covariance of power spectrum bandpowers (i.e. power spectra that have been binned in multipole space), which are subsequently used to obtain mock parameter constraints. The approximation is described in that section. However, the results provided in Sections 5.3 and 5.4.1 were obtained using the full (i.e. per-multipole) connected non-Gaussian covariance matrix for a single redshift bin. This took 36 days to calculate on 55 CPUs for 32 million elements.⁴ By contrast, the equivalent Gaussian and super-sample covariance matrices each took around an hour to calculate on 12 CPUs.

5.3. Comparison to simulations

5.3.1. Method

The publicly available⁵ full-sky simulated spin-2 shear maps of Takahashi et al. (2017) were used, which were produced by ray tracing through cosmological N-body simulations. The simulations use a maximum box size of $6300 h^{-1} \text{Mpc}$. These simulated maps are quite versatile, not only because they cover the full sky, but also because they describe the underlying shear field with no shape noise or shot noise (which could be added if required, but is not added for this work). A total of 108 realisations were performed, which is a relatively small number considering that very large numbers of realisations may be required for full convergence of a simulated covariance (Blot et al. 2016), but the results show that a useful comparison between theory and simulations is still possible. In addition, finite box simulations will necessarily underestimate the super-sample covariance (Hamilton, Rimes & Scoccimarro 2006; Li, Hu & Takada 2014), but a significant deficit is not detected. Further details and validation of the simulations can be found in Takahashi et al. (2017), and density maps, halo catalogues and lensed cosmic microwave background maps are also available at the same URL.

The shear maps are provided at 38 redshift slices from $z = 0.05$ to 5.3. For each realisation, these were combined into five redshift bins, following a Gaussian redshift distribution

⁴ This number comes from a data vector running from $\ell_{\min} = 2$ to $\ell_{\max} = 8000$, which has a length of $n = 8000 - 2 + 1 = 7999$, leading to a number of unique covariance elements equal to $n(n+1)/2 = 31\,996\,000$.

⁵ http://cosmo.phys.hirosaki-u.ac.jp/takahashi/allsky_raytracing

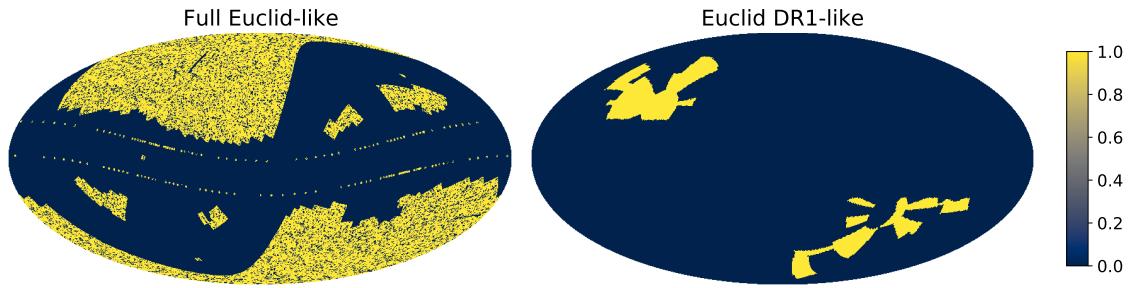


Figure 5.1. Full *Euclid*-like and *Euclid* DR1-like masks, which are used in Sections 5.3 and 5.4 to quantify the effects of different masks on the power spectrum covariance.

centred at $z = 0.65, 0.95, 1.25, 1.55, 1.85$ with a standard deviation of 0.3. The combined shear map for each redshift bin was formed as a weighted average over all 38 slices, with the weights given by the probability density of a Gaussian distribution with the appropriate mean and standard deviation. This choice of redshift distribution is discussed below, in Section 5.3.1.1. Three copies were then taken: one full-sky with no mask, one with a full *Euclid*-like mask including the survey footprint and a bright star mask ($f_{\text{sky}} = 0.31$), and one with a *Euclid* DR1-like footprint but no bright star mask ($f_{\text{sky}} = 0.06$). The full *Euclid*-like and *Euclid* DR1-like masks are shown in Figure 5.1. These masks approximate the coverage of the Euclid Wide Survey at different stages but do not exactly correspond to what will be observed, which is described in Euclid Collaboration: Scaramella et al. (2022). It is also assumed that the masks are uncorrelated with the signal, which may not be the case in practice (e.g. Fabbian et al. 2021). Finally, the `healpy` (Zonca et al. 2019) interface to the `HEALPix` (Górski et al. 2005) software was used to measure the spin-2 shear power spectra for each realisation. The comparisons shown in this section are for the E -mode auto-power in the lowest redshift bin.

For the theoretical covariance components described in Section 5.2, the same cosmology and redshift distribution as the simulations were used. A maximum multipole of $\ell_{\text{max}} = 8000$ was used in intermediate calculations to fully account for all relevant mode coupling, but the comparison was limited to $\ell \leq 3000$ because the $n_{\text{side}} = 4096$ maps that were used experience distortion from limited angular resolution above this point, as documented in Takahashi et al. (2017). Higher resolution maps with $n_{\text{side}} = 8192$ are also available, so the ℓ range could in principle be extended, albeit with significantly increased computational requirements.

5.3.1.1. Choice of redshift distribution

The choice of redshift distribution used here—five Gaussian bins, centred on $z = 0.65, 0.95, 1.25, 1.55, 1.85$ with a standard deviation of 0.3—was made for simplicity, with the relatively low number of redshift bins (five) also chosen for computational efficiency. Since there is freedom to enforce agreement between the redshift distributions in the simulations

and theory, there is little additional value in choosing a more complicated, more realistic distribution. Future cosmological analyses of real *Euclid* data are likely to use a larger number of bins with less overlap than is used here (e.g. Euclid Collaboration: Pocino et al. 2021). There is no reason that this will affect the results presented in this chapter, although it should be noted that marginalising over nuisance parameters describing photometric redshift uncertainties will reduce the importance of all cosmological contributions to the covariance.

5.3.2. Results

Figure 5.2 shows correlation matrices for the simulated covariance compared to the total theoretical covariance for each mask, as well as the individual components of the theory covariance. There appears to be good agreement between the simulated and total theoretical correlation matrices for all three masks. The relative contributions from the three covariance components are discussed in Section 5.4.

Figure 5.3 shows a detailed comparison of certain diagonals of the covariance matrix. For the main diagonal ($\Delta\ell = 0$), the variance divided by the square of the power spectrum is shown, to remove any effects coming from disagreement in the power spectrum between the simulations and theory, which is not the focus of this work. For the off-diagonals ($\Delta\ell = 2, 10, 100$) the correlation is shown. In all panels the simulated line is a rolling average over 50 ℓ s and the shaded region is the standard deviation over this range. For the full-sky and full *Euclid*-like masks, excellent agreement is observed between the theory and simulations. For the more extreme *Euclid* DR1-like mask there is a slightly worse, but still generally good, level of agreement. In particular, the super-sample covariance component is clearly correctly increasing with the severity of the sky cut to match the additional correlation found in the simulations. The relative sizes and importance of the three theory contributions are discussed further in Section 5.4. We may conclude from Figures 5.2 and 5.3 that CosmoLike’s non-Gaussian covariance calculations appear to be suitably accurate, to the degree that can be assessed using these simulations.

This comparison was also repeated for purely Gaussian fields, which were simulated using `healpy`. The results are shown in Figure 5.4, where an excellent level of agreement is observed between the Gaussian field simulations and the prediction from the improved NKA, which is a significant improvement over the standard NKA (not shown).

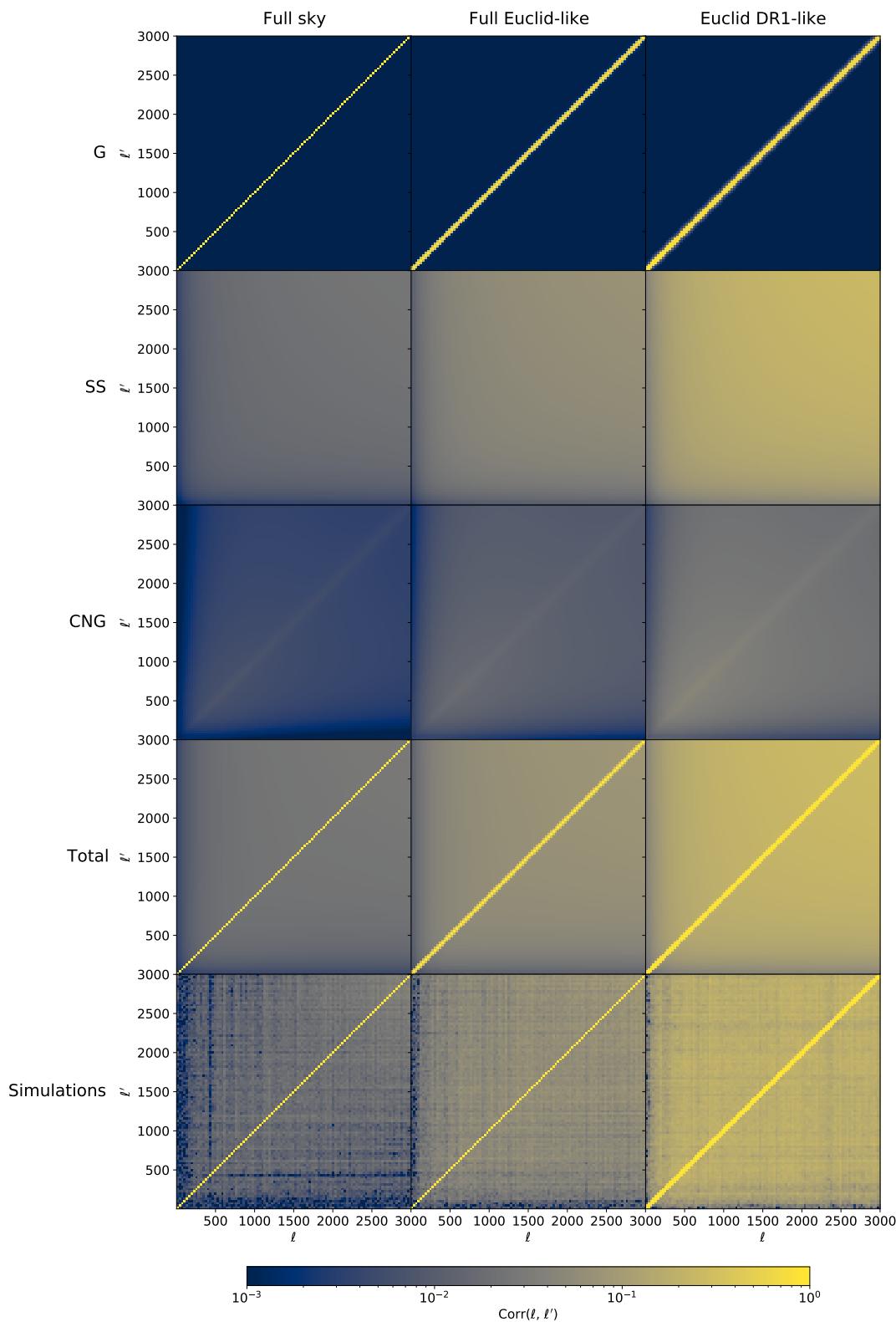


Figure 5.2. Correlation matrices for the simulated covariance compared to the total theoretical covariance for each mask and for the individual components of the theory covariance: Gaussian (G), super-sample (SS), and connected non-Gaussian (CNG). The covariance shown here is for the shear E -mode power spectrum in the lowest redshift bin, without shape noise.

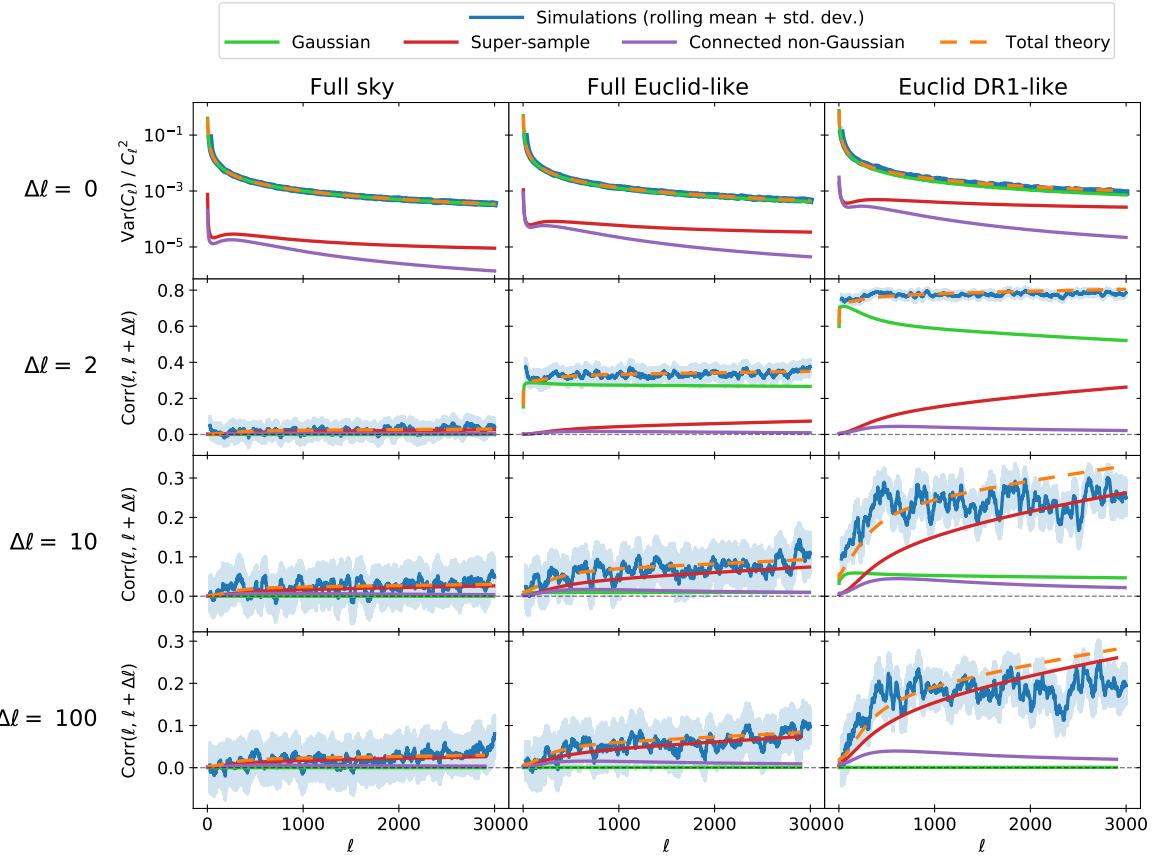


Figure 5.3. Comparison between the covariance predicted by theory and measured from simulations, for the three masks. The top row shows the variance divided by the power spectrum squared, and the lower three rows show correlation. In all panels the simulated line is a rolling average over 50 ℓ s, and the shaded region is the standard deviation over this range. The covariance shown here is for the shear E -mode power spectrum in the lowest redshift bin, without shape noise.

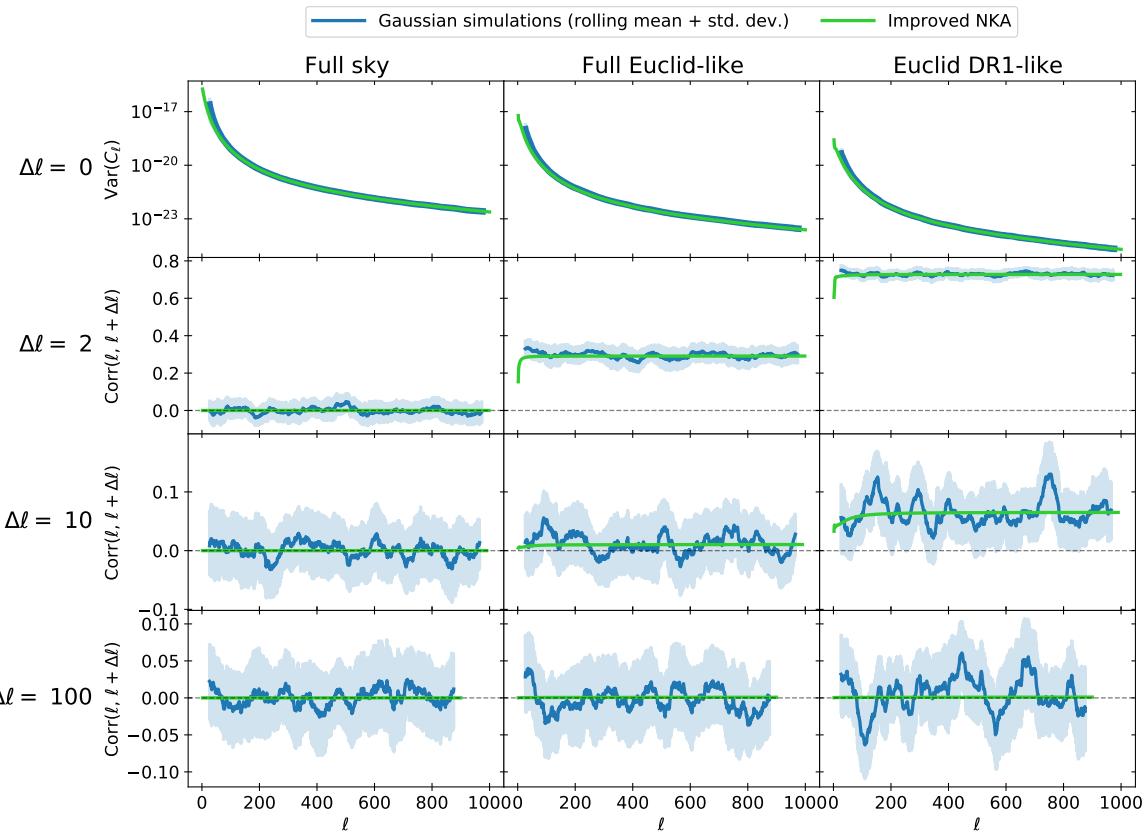


Figure 5.4. Comparison between covariance measured from Gaussian field simulations and predicted using the improved NKA method, for the three masks. In all panels the simulated line is a rolling average over 50 ℓ s, and the shaded region is the standard deviation over this range. No shape noise is included.

5.4. Importance of covariance components and dependence on mask

This section addresses the size and importance of the different components of the cosmic shear pseudo- C_ℓ covariance, and how these properties depend on the mask.

5.4.1. Relative sizes of components

5.4.1.1. Without shape noise

Let us first consider the case without shape noise, which has already been shown in Figures 5.2 and 5.3. It can be seen from the full correlation matrices plotted in Figure 5.2 that the main diagonal of the matrix is dominated by the Gaussian component, which is purely diagonal in the full-sky case and visibly broadens slightly as the sky cut is increased. The super-sample covariance is the dominant off-diagonal component, particularly at higher ℓ but extending down visibly even to $\ell < 500$ in the case of the most extreme *Euclid* DR1-like mask. The connected non-Gaussian contribution is barely visible on the colour scale, other than for the *Euclid* DR1-like mask at low ℓ .

A more detailed comparison of the relative sizes of the different components is possible with the selected diagonals shown in Figure 5.3. Again it is apparent that the main diagonal ($\Delta\ell = 0$) is dominated by the Gaussian component, but the extent to which this is the case is reduced as the sky cut is increased, as the contribution increases from both non-Gaussian components. Moving away from the main diagonal, at $\Delta\ell = 2$ the Gaussian is still the largest component (except on the full sky, where its contribution is purely diagonal), but by $\Delta\ell = 10$ the super-sample component is dominant and increasing towards higher ℓ . It is clear that the super-sample covariance contribution increases with a more severe sky cut, though notably it is still visibly non-zero even for full-sky observations. The connected non-Gaussian component is the subdominant non-Gaussian contribution at all values of ℓ and $\Delta\ell$ for all masks.

5.4.1.2. With shape noise

Figure 5.5 shows an equivalent comparison of the sizes of the theoretical covariance components with shape noise included. Gaussian shape noise is assumed, which—as introduced in Chapter 2—is included as a contribution to the power spectrum,

$$C_\ell \rightarrow C_\ell + N_\ell; \quad (5.8)$$

$$N_\ell = \frac{\sigma_e^2}{N_i}, \quad (5.9)$$

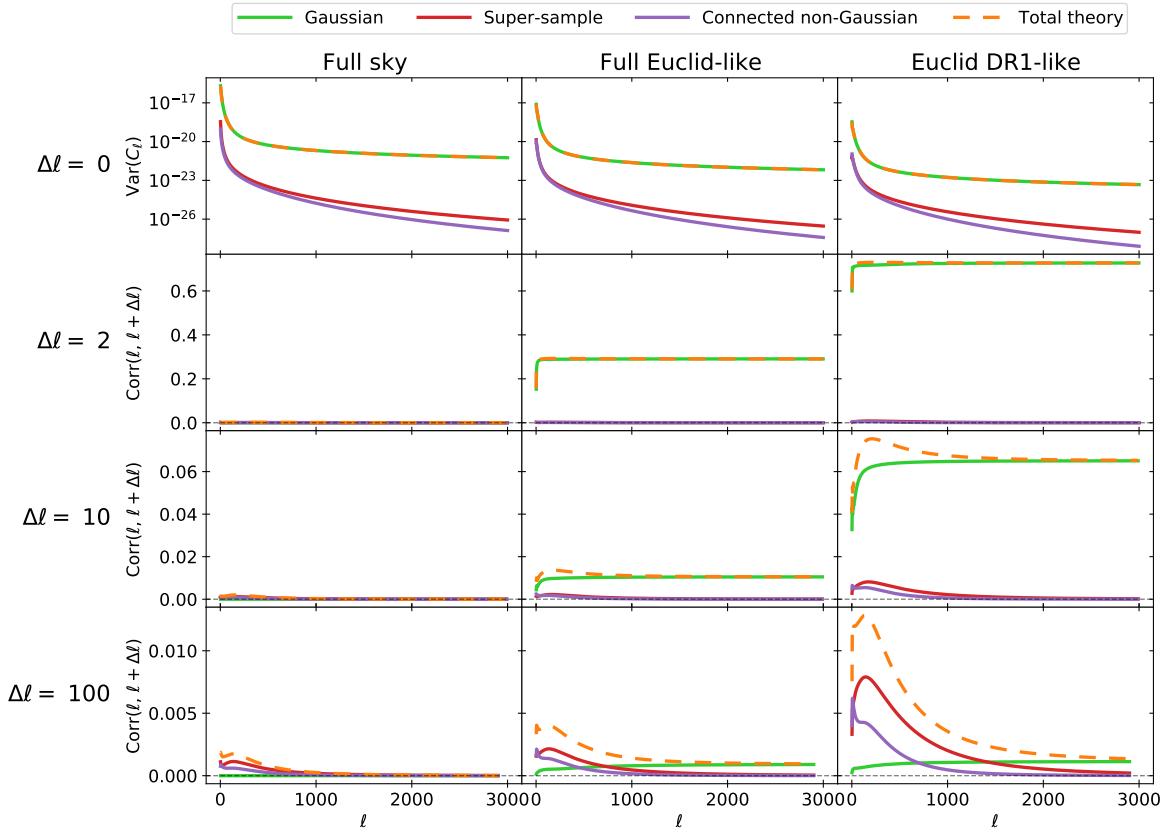


Figure 5.5. Comparison between contributions to the theoretical covariance for the three masks, with shape noise included following Equation (5.9). The top row shows the variance, and the lower three rows show correlation.

where σ_ϵ is the intrinsic galaxy shape dispersion per component and N_i is the number of galaxies per steradian per redshift bin. A *Euclid*-like galaxy number density of $30 / \text{arcmin}^2$, equally split over five redshift bins, was assumed, along with a value of $\sigma_\epsilon = 0.3$.

With shape noise included, quite different behaviour to the no-noise case is found. The Gaussian-dominated main diagonal is substantially increased, especially at higher ℓ , resulting in non-Gaussian off-diagonal correlations being significantly suppressed. The result is that the Gaussian component is dominant at all ℓ as far away from the main diagonal as $\Delta\ell = 10$. By $\Delta\ell = 100$, the Gaussian component is no longer dominant at lower ℓ , but continues to account for the largest contribution at higher ℓ : above $\ell \sim 1000$ for the full *Euclid*-like mask and $\ell \sim 1500$ for the *Euclid* DR1-like mask. This suggests that once shape noise is included, the Gaussian component is more important than the no-noise results in Figure 5.3 suggest.

5.4.2. Importance for parameter constraints

While the relative size of the different covariance components studied in Section 5.4.1 offers interesting insight into their behaviour, it says relatively little about the actual importance of each component. In particular, it is unclear to what extent the dominance of the Gaussian component on and close to the main diagonal is offset by its sub-dominance farther away from the main diagonal. In this section, to gain more insight into this, mock parameter constraints are produced. Shape noise is included, following Equation (5.9).

Here five redshift bins were used, including all auto- and cross-spectra (E -modes only), giving 15 power spectra in total. Scales up to $\ell_{\max} = 5000$ were included. The full data vector for this setup would have $n = 75\,000$ elements (15×5000), which gives $n(n+1)/2 = 2.8$ billion unique covariance elements. Due to the time needed to evaluate the projected matter trispectrum, it would be unfeasible to calculate the connected non-Gaussian contribution in full. As a result, an angular binning approach was taken, with 12 logarithmically spaced bandpowers, and an approximation was used to obtain the connected non-Gaussian bandpower covariance from a more modest number of per- ℓ covariance calculations. This approximation is described and validated in Section 5.4.2.1. The Gaussian and super-sample covariance components were calculated in full, using scales up to $\ell_{\max} = 8000$ for intermediate calculations, before being binned into bandpowers as

$$P_b = \sum_{\ell} \mathbf{P}_{b\ell} C_{\ell}, \quad (5.10)$$

where \mathbf{P} is the bandpower binning matrix whose elements are given by

$$\mathbf{P}_{b\ell} = \begin{cases} \frac{\ell(\ell+1)}{2\pi} \left[\ell_{\min}^{b+1} - \ell_{\min}^b \right]^{-1} & \text{for } \ell_{\min}^b \leq \ell < \ell_{\min}^{b+1}; \\ 0 & \text{otherwise,} \end{cases} \quad (5.11)$$

where ℓ_{\min}^b is the lower edge of bin b .

A mock observation was obtained by sampling from a Gaussian likelihood with the total covariance. The input mean was the fiducial theory power spectra, plus noise for auto-spectra given by Equation (5.9), mixed using the mixing matrix obtained using NaMaster, then binned following Equation (5.10). This random sampling process replicates the randomness of cosmic variance that is present in a real observation, and means that—as with real data—the resulting posterior distributions are not centred on the ‘true’ input parameters. Checks similar to those shown in Section 4.4 confirmed that bandpowers measured from the Takahashi et al. (2017) simulations used in Section 5.3 are no more non-Gaussian than those measured from Gaussian field simulations, and therefore since a Gaussian likelihood was shown to be sufficiently accurate for Gaussian fields in Chapter 4, it is a suitable choice here.

Parameter constraints were obtained by iterating over two-parameter grids produced using CosmoSIS (Zuntz et al. 2015), following the pipeline described in Section 4.2.2.1.⁶ All other parameters were held fixed. At each point in parameter space, theory bandpowers—calculated in the same way as the input mean to the observation described above—were compared to the observed bandpowers using a Gaussian likelihood with different combinations of covariance components. All combinations necessarily include the Gaussian component, since this on its own is a valid positive definite covariance matrix, unlike the super-sample and connected non-Gaussian components.

5.4.2.1. Connected non-Gaussian approximation

As noted above, it is impractical to calculate the connected non-Gaussian component for all 2.8 billion unique elements of the full covariance matrix. Instead, an approximation was used to directly obtain its contribution to the bandpower covariance. This approximation was designed to mimic two effects: the mixing of power by the survey mask, and the binning of individual multipoles into bandpowers. In this analysis, both of these processes are cosmology-independent: each shear field uses the same mask, and all power spectra use the same binning scheme. As a result, both processes should have approximately the same effect on every power spectrum, and consequently also every covariance block. Therefore, the approximation made here uses two sets of weights—one to mimic binning and the other to mimic mixing—which were calibrated for the covariance of the shear auto-power spectrum in the lowest redshift bin and then applied to all further blocks.

First, the connected non-Gaussian component was evaluated in full for a single covariance element per bandpower pair, for all combinations of power spectra. This was chosen to be for the weighted average ℓ in each bandpower, with the weights given by \mathbf{P}_{bl} (Equation 5.11), rounded to the nearest integer. This vastly reduced the number of projected trispectrum calculations, to 16 290.⁷ The result was then re-weighted using the weights calibrated using the covariance of the shear auto-power spectrum in the lowest redshift bin, which was calculated in full for the previous sections.

The weighting can be understood as a two-step process. First, a ‘binning’ weighting was applied, designed to mimic the effect of taking the full unbinned covariance matrix and binning it into bandpowers. Then a ‘mixing’ weighting was applied, designed to mimic the effect of the mixing matrix. In both cases, the weights were obtained by carrying out the process in full for the shear auto-power spectrum in the lowest redshift bin.

This can be illustrated using equations as follows. For the first block (covariance of shear auto-power in the lowest redshift bin), the following procedure was used to transform the

⁶ The pipeline includes the CAMB (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012) and Halofit-Takahashi (Smith et al. 2003; Takahashi et al. 2012) modules.

⁷ This number comes from a reduced data vector of 12 bandpowers and 15 power spectra, giving a data vector of length $n = 12 \times 15 = 180$ and a number of unique covariance elements of $n(n + 1)/2 = 16\,290$.

full unbinned covariance $\mathbf{Cov}_{\text{unbinned}}$ into a final binned and mixed block $\mathbf{Cov}_{\text{mixed}}$, via a binned and unmixed stage $\mathbf{Cov}_{\text{binned}}$:

$$\mathbf{Cov}_{\text{binned}} = \mathbf{P} \mathbf{Cov}_{\text{unbinned}} \mathbf{P}^T; \quad (5.12)$$

$$\mathbf{Cov}_{\text{mixed}} = \mathbf{M} \mathbf{Cov}_{\text{binned}} \mathbf{M}^T, \quad (5.13)$$

where \mathbf{P} and \mathbf{M} are the bandpower binning and pseudo- C_ℓ mixing matrices, respectively. Elements were selected from $\mathbf{Cov}_{\text{unbinned}}$ corresponding to the weighted average ℓ within each bandpower to give $\mathbf{Cov}_{\text{sampled}}$. The matrices of binning weights \mathbf{w}_{bin} and mixing weights \mathbf{w}_{mix} were then calculated as

$$\mathbf{w}_{\text{bin}} = \frac{\mathbf{Cov}_{\text{binned}}}{\mathbf{Cov}_{\text{sampled}}} \quad (\text{elementwise}); \quad (5.14)$$

$$\mathbf{w}_{\text{mix}} = \frac{\mathbf{Cov}_{\text{mixed}}}{\mathbf{Cov}_{\text{binned}}} \quad (\text{elementwise}). \quad (5.15)$$

Finally, the sampled covariance blocks $\mathbf{Cov}_{\text{sampled}}$ were calculated for every block in the full covariance matrix, and transformed to give approximate binned and mixed covariance blocks as

$$\mathbf{Cov}_{\text{binned_approx}} = \mathbf{w}_{\text{bin}} * \mathbf{Cov}_{\text{sampled}} \quad (\text{elementwise}); \quad (5.16)$$

$$\mathbf{Cov}_{\text{mixed_approx}} = \mathbf{w}_{\text{mix}} * \mathbf{Cov}_{\text{binned_approx}} \quad (\text{elementwise}). \quad (5.17)$$

While this two-step weighting process could be equivalently formulated as a single step, separating the effect of the binning and mixing approximations allows for additional insight into their respective effects.

These approximations were validated by carrying out an equivalent process for the super-sample covariance matrix and comparing the results to those obtained using the full correct treatment. Histograms of the ratios between the approximate and exact covariance for each step, for all elements of the bandpower covariance across all redshift bins, are shown in Figure 5.6. Each step introduces a bias of order per cent on average (although conveniently in opposite directions) with a spread of a few per cent. This is sufficiently accurate for the purposes of this work, especially considering that the connected non-Gaussian component is the smallest of the three, but nevertheless this small potential error should be borne in mind when interpreting the results. Since the super-sample and connected non-Gaussian covariance contributions were shown to be similarly smooth in Section 5.3, the fact that this approximation works well for the former suggests that it should too for the latter.

5.4.2.2. Results

Two-parameter constraints for different combinations of covariance components are shown in Figure 5.7. The top row shows dark energy equation of state parameters (w_0, w_a), where

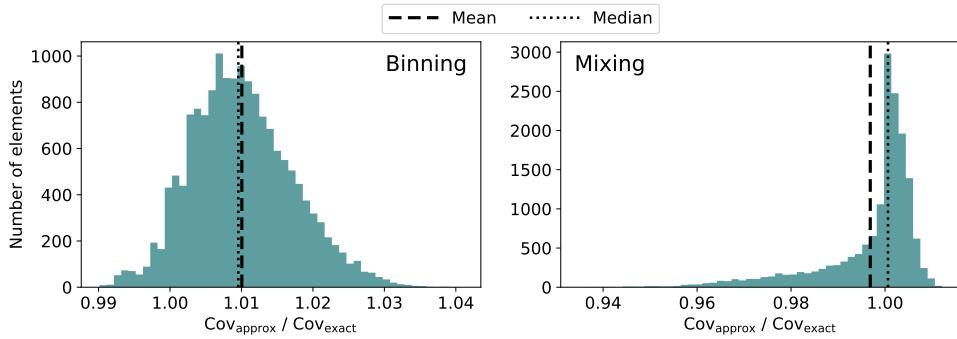


Figure 5.6. Validation of the connected non-Gaussian approximation used to obtain the mock parameter constraints in Section 5.4.2, which is described in Section 5.4.2.1. Histograms of the ratio of the approximate to exact covariance are shown, for the ‘binning’ (left) and ‘mixing’ (right) steps, for all elements of the bandpower covariance matrix across all redshift bins, measured using the super-sample covariance. The results in all other sections are obtained using the connected non-Gaussian component calculated in full.

$w(a) = w_0 + w_a(1-a)$. The bottom row shows the matter density Ω_m and the amplitude of the matter power spectrum at $z=0$ on the scale of $8 h^{-1} \text{ Mpc}$, σ_8 . The three columns are for the three different masks. In each panel, all parameters other than the two shown are held fixed. Only the 1 and 3σ credible regions are marked, which respectively contain the highest 68.3 and 99.7 per cent of the posterior probability mass. The relative areas of the 3σ credible region are listed for each combination of parameters and mask in Table 5.1.

The Gaussian contribution (G) alone only covers 30–38 per cent of the full 3σ region for (w_0, w_a) and 37–49 per cent for (Ω_m, σ_8) . The Gaussian and connected non-Gaussian components combined (G + CNG) cover 51–63 and 54–70 per cent for (w_0, w_a) and (Ω_m, σ_8) , respectively, while the Gaussian and super-sample components combined (G + SS) cover 82–84 and 90–92 per cent. These results are broadly in line with the single-parameter error bar ratios obtained in Barreira, Krause & Schmidt (2018b).

There is some amount of apparent mask dependence: as the sky cut is increased, the relative area of G + SS sees a very small increase (by 2–3 per cent) whereas G + CNG and G see larger decreases (by 16–22 and 8–12 per cent, respectively). This is consistent with the expectation that super-sample covariance should become more important as the sky is cut further, since this excludes more modes from the survey.

There are also some small shifts in the posterior means between the different composite covariance results for a given mask, despite the fact that the mean of the Gaussian likelihood used in each case is identical and only the covariance differs. This demonstrates how an incorrect covariance leads to an incorrect weighting of the random scatter present in the data due to cosmic variance, and therefore to posterior constraints having not only the wrong size but an erroneous position too, although this is a small effect.

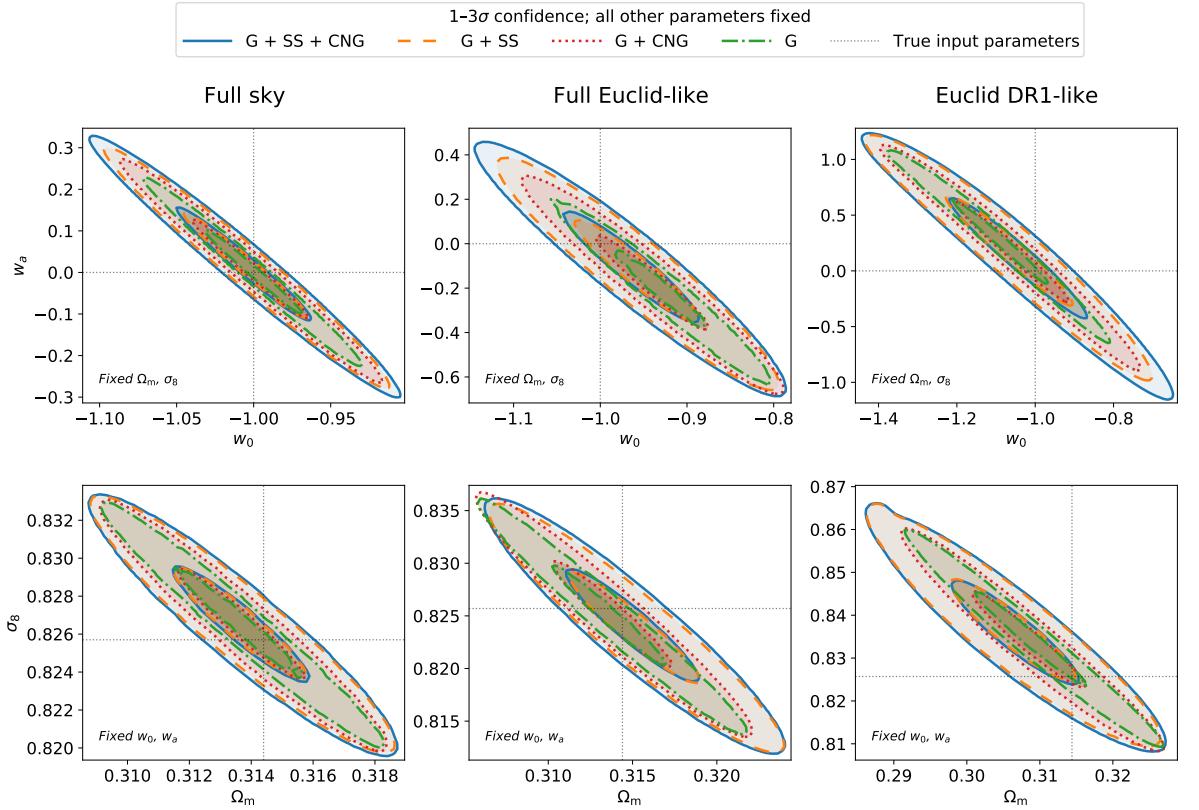


Figure 5.7. Two-parameter constraints for different masks and different combinations of covariance contributions: Gaussian (G), super-sample (SS), and connected non-Gaussian (CNG). Shape noise is included, following Equation (5.9). In each panel, all parameters other than the two shown are held fixed. Only the 1 and 3σ credible regions are marked, which respectively contain the highest 68.3 and 99.7 per cent of the posterior probability mass. The relative areas of each 3σ credible region are listed in Table 5.1. Note that the axis ranges differ between panels.

Table 5.1. Relative areas of 3σ credible regions in Figure 5.7.

Parameters	Mask	Relative area of 3σ credible region (%)			
		G + SS + CNG	G + SS	G + CNG	G
(w_0, w_a)	Full sky	100	82	63	38
	Full <i>Euclid</i> -like mask	100	84	61	36
	<i>Euclid</i> DR1-like mask	100	85	51	30
(Ω_m, σ_8)	Full sky	100	90	70	49
	Full <i>Euclid</i> -like mask	100	90	66	45
	<i>Euclid</i> DR1-like mask	100	92	54	37

The table shows the relative area of the 3σ credible region in Figure 5.7 for each combination of parameters and mask and for different combinations of covariance components: Gaussian (G), super-sample (SS), and connected non-Gaussian (CNG). The relative 1σ areas are similar to the 3σ areas.

5.4.2.3. Effect of marginalisation over additional parameters

Motivation to explore the impact of marginalisation over additional parameters on the relative importance of each covariance component is provided by the results of Barreira, Krause & Schmidt (2018b). In that paper, the authors produced mock parameter constraints with different covariance contributions included, both for a single parameter at a time with all others fixed and for five parameters with all five allowed to vary simultaneously. For the latter case, they display marginalised two-parameter constraints (Barreira, Krause & Schmidt 2018b, their Figure 3). For (w_0, w_a) in particular (and to a lesser extent with some other parameter pairs), the 2σ credible region obtained with the Gaussian covariance alone appears roughly the same as that for the total covariance, to within the sampling noise. This is in contrast to their single-parameter constraints, for which the Gaussian covariance only produced ~ 50 per cent of the 1σ uncertainty on w_0 obtained using the full covariance (see Figure 2 of Barreira, Krause & Schmidt 2018b). This raises the question of whether marginalisation might reduce the differences between the constraints obtained using different combinations of covariance components.

Here, this question is investigated by performing a three-parameter likelihood analysis using the full *Euclid*-like mask over (w_0, w_a, Ω_m) , with all other parameters still held fixed. One- and two-parameter marginalised constraints are shown in Figure 5.8. Table 5.2 compares the relative areas and widths of one- and two-parameter 3σ credible regions before and after marginalisation over a third parameter.

While the relative areas in Figure 5.8 appear qualitatively similar to those in Figure 5.7, there are in fact some substantial quantitative differences, as shown by the values in Table 5.2. In particular, there is almost a doubling in the area of the constraints on (w_0, w_a) from the Gaussian covariance only (G) relative to the total covariance (G + SS + CNG)—from 36 to 69 per cent—when marginalising over Ω_m rather than holding it fixed. A similar but smaller increase is seen for the other subsets (G + SS, G + CNG) of the total covariance for the same parameters.

For constraints on w_a alone, the width of the 3σ credible region for G relative to G + SS + CNG increases slightly when marginalising over both w_0 and Ω_m compared to only marginalising over w_0 , and a similar slight increase is seen for G + SS and G + CNG. However, for constraints on w_0 , there is a small decrease in relative widths when marginalising over both w_a and Ω_m rather than only w_a . One reason for this difference in behaviour between w_0 and w_a may be that—as seen in Figure 5.8—there is clearly a much stronger correlation between w_a and Ω_m than between w_0 and Ω_m . Marginalisation over a strongly correlated parameter should broaden constraints more than marginalisation over a more weakly correlated parameter (indeed, marginalisation over a truly independent parameter should have no effect at all), but it is not obvious that this should change the ratio of relative areas rather than simply broadening all constraints by the same factor. Regardless of the origin of this behaviour, it does appear to be the case that marginalisation

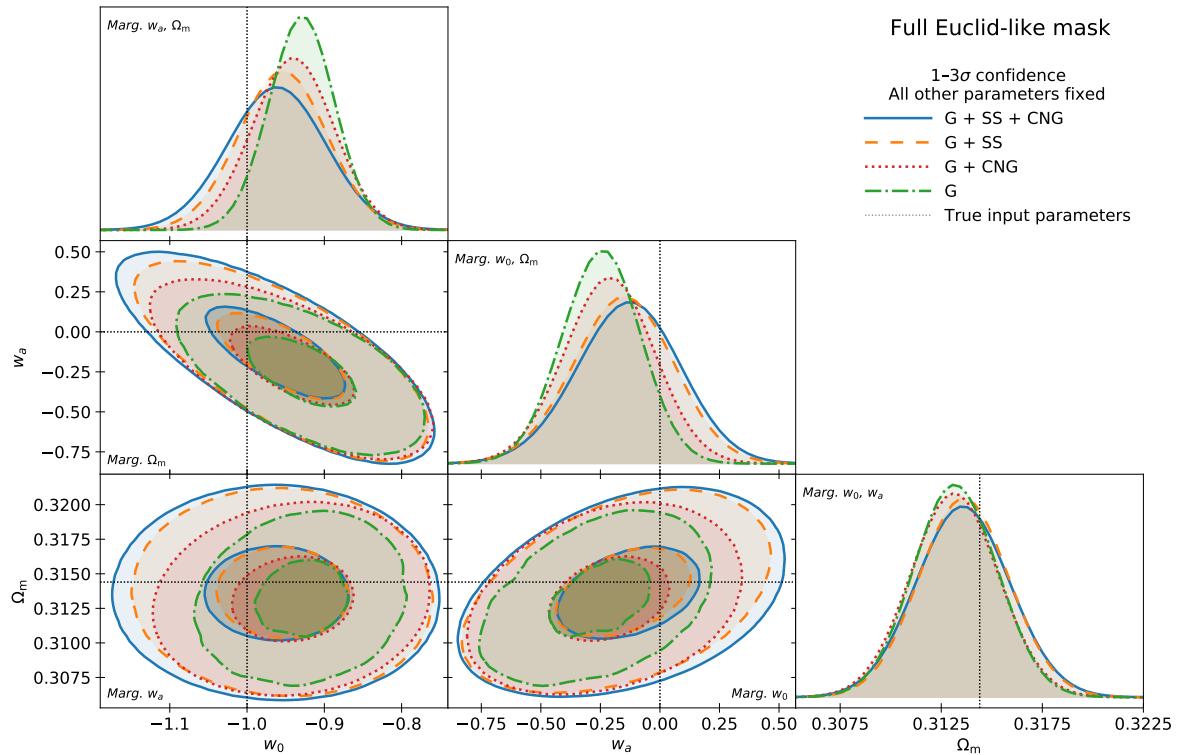


Figure 5.8. Two- and one-parameter marginalised constraints obtained from a joint three-parameter analysis of (w_0, w_a, Ω_m) for the full *Euclid*-like mask, including different combinations of covariance contributions: Gaussian (G), super-sample (SS), and connected non-Gaussian (CNG). Shape noise is included, following Equation (5.9). The constraints in each panel have been obtained by marginalising over one or two parameters in the joint three-parameter posterior; for example, the panel marked ‘Marg. w_a, Ω_m ’ has been marginalised over w_a and Ω_m . All other parameters are held fixed. Only the 1 and 3σ credible regions are marked.

Table 5.2. Impact of marginalisation on 3σ credible regions.

Parameter(s)	Marginalised over	Relative area or width of 3σ credible region (%)			
		G + SS + CNG	G + SS	G + CNG	G
(w_0, w_a)	—	100	84	61	36
	Ω_m	100	89	84	69
w_0	w_a	100	91	88	72
	(w_a, Ω_m)	100	90	83	69
w_a	w_0	100	88	84	70
	(w_0, Ω_m)	100	96	85	74

The table shows the relative areas (for two-parameter constraints) and widths (for one-parameter constraints) of the 3σ credible regions obtained using different combinations of covariance contributions, Gaussian (G), super-sample (SS), and connected non-Gaussian (CNG), for the full *Euclid*-like mask. Each row contains two sub-rows: the top sub-row is based on a two-parameter fit, which is marginalised over zero or one parameters; the bottom sub-row is based on a three-parameter fit, which is marginalised over one or two parameters. The relative 1σ areas are similar to the 3σ areas.

over additional parameters—particularly those with which the constrained parameters are correlated—affects the relative importance of the difference covariance contributions. This is in agreement with the findings of Barreira, Krause & Schmidt (2018b).

5.5. Conclusions

As the era of next-generation weak lensing surveys such as *Euclid* rapidly approaches, it is increasingly important to understand the properties of all steps of an analysis pipeline, including the covariance used in the likelihood. Section 5.2 has described how existing publicly available codes can be used in combination to calculate the full covariance matrix of cosmic shear pseudo- C_ℓ estimates, including the full details of an arbitrary mask. It has been further shown in Section 5.3 that existing simulations can be used to verify the accuracy of a theoretical covariance, which found a high degree of agreement and consistency between theory and simulations. This agreement persists for different masks, showing that the theoretical covariance contributions correctly account for both the cut-sky mode coupling that is inherent to the pseudo- C_ℓ method and the non-Gaussian mode coupling, including additional cut-sky super-sample covariance.

This is encouraging for the use of pseudo- C_ℓ estimators in weak lensing, whose convenience and speed make them an attractive choice of analysis framework for future surveys. However, an outstanding challenge with such estimators is the need to understand their statistical properties sufficiently well such that they can be used to deliver reliable cos-

mological constraints to the precision and accuracy needed by future high-precision weak lensing surveys. This challenge has now to a large degree been addressed since it is now known that not only is a Gaussian likelihood sufficient (Chapter 4), but a full covariance can be evaluated and validated using the methods shown in this chapter.

The results in this chapter have demonstrated that it is essential to include the non-Gaussian contributions to the covariance, even though cut-sky mode coupling means that the Gaussian covariance component dominates off-diagonal modes close to the main diagonal. The relative size and importance of the Gaussian component increases when including shape noise, but it has been shown in Section 5.4 that only including the Gaussian component in parameter inference can lead to an underestimation of uncertainties by up to 70 per cent. The dominant non-Gaussian covariance component is the super-sample covariance, but neglecting the subdominant connected non-Gaussian covariance component can still lead to uncertainty underestimation on the scale of 10–20 per cent. In addition, neglecting some covariance contributions can lead to biases in the position of posterior parameter constraints as well as their size. However, a real cosmological analysis will require marginalisation over many nuisance parameters, which will decrease the relative importance of all cosmological contributions to the covariance, so these values should be taken as upper limits on the importance of each component. Perhaps for this reason it was found in the analysis of DES Year 3 data in Friedrich et al. (2021) that the connected non-Gaussian contribution could be entirely neglected, but the results of this work suggest that this conclusion cannot be automatically extended to a *Euclid*-like survey. However, this need not be an inconvenience, since approximations of the kind described in Section 5.4.2.1 can be used to obtain the connected non-Gaussian component in a manageable amount of time with a loss of accuracy of only a few per cent on an already subdominant term. Finally, this chapter has shown that marginalisation over additional cosmological parameters may have a substantial effect on the relative importance of the different covariance components. We may conclude from this that it is important to take all marginalisation into account when, for example, determining the required accuracy of theoretical results for a particular science goal or producing forecasts of parameter constraints. The consistency of these results with those of Barreira, Krause & Schmidt (2018b) implies that cut-sky mode coupling has relatively little impact on the respective importance of the covariance components.

Chapter 6

Dependence of cosmological parameter constraints on angular binning of weak lensing two-point statistics

6.1. Introduction

As introduced in Chapter 2, two-point statistics of weak lensing shear and galaxy number overdensity may be calculated in spherical harmonic space to estimate power spectra, or in real (also called configuration) space to estimate correlation functions. In either case, it is usually necessary to apply an angular binning in ℓ or θ . For the power spectrum, this is not strictly necessary in principle, provided the effect of the survey mask is forward modelled rather than removed with a mixing matrix inversion (see the discussion of the pseudo- C_ℓ method in Section 2.3). However, a realistic 3×2 pt analysis setup with 10 tomographic redshift bins and a scale cut of $\ell_{\max} = 5000$ would have a data vector with over a million elements if no angular binning was applied, which would be highly impractical. Furthermore, the smooth nature of weak lensing power spectra means that retaining a perfect angular resolution is probably unnecessary. For the correlation function, on the other hand, it is intrinsically necessary to bin in θ , since galaxy pair separation is a continuous quantity sampled at particular points determined by the set of galaxies in the survey.

Varying numbers of angular bins have been used in recent weak lensing analyses. The Dark Energy Survey Year 3 analysis in DES Collaboration et al. (2022) used 20 correlation function bins from 2.5 to 250 arcmin, while the KiDS-1000 analysis used 9 bins from 0.5 to 300 arcmin, and 8 bandpowers from $\ell = 100$ to 1500 (Joachimi et al. 2021; Heymans et al. 2021; Asgari et al. 2021). The Hyper Suprime-Cam Year 1 pseudo- C_ℓ analysis in Hikage et al. (2019) used 15 bandpowers from $\ell = 60$ to 6500, though only 6 from $\ell = 300$ to 1900 were retained in the cosmological analysis.

This chapter explores how many angular bins are necessary to retain a sufficient amount of constraining power in cosmological parameters. Specifically, it studies how the relative size of posterior uncertainties depends on the number of angular bins, both for the power spectrum and correlation function. Since angular binning will be essential in practice for upcoming Stage IV weak lensing surveys, it is vital to understand the relationship between this binning and cosmological parameter constraints in order to strike the optimal balance between computational viability and scientific value.

The main analysis of this chapter is contained within Section 6.2, which studies how the size of parameter posterior uncertainties depends on the number of angular bins. In Section 6.3 a number of additional effects that may affect these results are examined. Conclusions are discussed in Section 6.4.

6.2. Dependence of posterior uncertainties on number of angular bins

This section contains the main analysis of the chapter. The method of measuring posterior uncertainties is described in Section 6.2.1, while the observed dependence on the number of angular bins is described in Section 6.2.2.

6.2.1. Measurement of posterior uncertainties

The posterior uncertainties in this section are measured directly from posterior distributions resulting from a full likelihood analysis. Later in the chapter, in Section 6.3, methods will be described and used which approximate these results without the need for a full likelihood analysis.

The different elements of the likelihood analysis will now each be described.

6.2.1.1. Modelling of two-point statistics

3×2 pt power spectra were generated using CosmoSIS (Zuntz et al. 2015), covering a two-dimensional grid of values of w_0 and w_a , and one-dimensional grids of five other cosmological parameters (Ω_m , Ω_b , σ_8 , n_s , h), with all other parameters held at fixed values in each case. The pipeline is the same as described in Section 4.2.2.1, with five Gaussian redshift bins centred on $z = 0.65, 0.95, 1.25, 1.55, 1.85$. Multipoles up to $\ell_{\max} = 5000$ were generated, but different scale cuts are explored in this chapter. Power spectra were binned into bandpowers following Equation (5.11).

Angular-bin-averaged correlation functions for galaxy position ξ_{NN} , shear ξ_{\pm} , and their cross correlation ξ_{NE} , were calculated directly from the unbinned power spectra using the following relations:

$$\xi_{NN}(\theta, \Delta\theta) = \sum_{\ell} \frac{2\ell+1}{4\pi} C_{\ell} \overline{d_{00}^{\ell}}(\theta, \Delta\theta); \quad (6.1)$$

$$\xi_{\pm}(\theta, \Delta\theta) = \sum_{\ell} \frac{2\ell+1}{4\pi} C_{\ell} \overline{d_{\pm 22}^{\ell}}(\theta, \Delta\theta); \quad (6.2)$$

$$\xi_{NE}(\theta, \Delta\theta) = \sum_{\ell} \frac{2\ell+1}{4\pi} C_{\ell} \overline{d_{20}^{\ell}}(\theta, \Delta\theta), \quad (6.3)$$

where $\overline{d_{m'm}^\ell}(\theta, \Delta\theta)$ is the bin-averaged Wigner small- d symbol. Following Fang et al. (2020) and Stebbins (1996), the small- d symbols may be written in terms of Legendre polynomials P_ℓ , associated Legendre polynomials of order 2 $P_\ell^{m=2}$,¹ and Stebbins's G symbol,

$$d_{00}^\ell(\theta) = P_\ell(\cos\theta); \quad (6.6)$$

$$d_{\pm 22}^\ell(\theta) = \frac{2}{\ell^2(\ell+1)^2} [G_{\ell,2}^+(\cos\theta) \pm G_{\ell,2}^-(\cos\theta)]; \quad (6.7)$$

$$d_{20}^\ell(\theta) = \frac{1}{\ell(\ell+1)} P_\ell^{m=2}(\cos\theta), \quad (6.8)$$

which may be integrated analytically to obtain (Friedrich et al. 2021; Fang et al. 2020)

$$\overline{P_\ell}(\theta, \Delta\theta) = \frac{1}{\cos\theta_2 - \cos\theta_1} \frac{1}{2\ell+1} [P_{\ell+1}(x) - P_{\ell-1}(x)]_{x=\cos\theta_1}^{x=\cos\theta_2}; \quad (6.9)$$

$$\begin{aligned} \overline{G_{\ell,2}^+ \pm G_{\ell,2}^-}(\theta, \Delta\theta) &= \frac{1}{\cos\theta_2 - \cos\theta_1} \left[-\frac{\ell(\ell-1)}{2} \left(\ell + \frac{2}{2\ell+1} \right) P_{\ell-1}(x) \right. \\ &\quad - \frac{\ell(\ell-1)(2-\ell)}{2} x P_\ell(x) + \frac{\ell(\ell-1)}{2\ell+1} P_{\ell+1}(x) \\ &\quad + (4-\ell) \frac{dP_\ell(x)}{dx} + (\ell+2) \left(x \frac{dP_{\ell-1}(x)}{dx} - P_{\ell-1}(x) \right) \\ &\quad \left. \pm 2(\ell-1) \left(x \frac{dP_\ell(x)}{dx} - P_\ell(x) \right) \mp 2(\ell+2) \frac{dP_{\ell-1}(x)}{dx} \right]_{x=\cos\theta_1}^{x=\cos\theta_2}; \end{aligned} \quad (6.10)$$

$$\begin{aligned} \overline{P_\ell^{m=2}}(\theta, \Delta\theta) &= \frac{1}{\cos\theta_2 - \cos\theta_1} \left[\left(\ell + \frac{2}{2\ell+1} \right) P_{\ell-1}(x) + (2-\ell) x P_\ell(x) \right. \\ &\quad \left. - \frac{2}{2\ell+1} P_{\ell+1}(x) \right]_{x=\cos\theta_1}^{x=\cos\theta_2}, \end{aligned} \quad (6.11)$$

where $\theta_1 = \theta$ and $\theta_2 = \theta + \Delta\theta$. The only symbols necessary to evaluate these equations are the Legendre polynomials and their derivatives, which are evaluated up to a required ℓ_{\max} using the SciPy Python library (Virtanen et al. 2020). These transforms assume a uniform distribution of galaxies, which corresponds to a distribution of galaxy pairs proportional to $\sin\theta$ (Friedrich et al. 2021). The angular bins are logarithmically spaced for both the power spectra and correlation functions.

Noise was added to the power spectra prior to binning, following Equations (4.8)–(4.9),

¹ Associated Legendre polynomials are defined as

$$P_\ell^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_\ell(x), \quad (6.4)$$

such that

$$P_\ell(x) \equiv P_\ell^{m=0}(x). \quad (6.5)$$

with a number density of $6/\text{arcmin}^2$ in each of five redshift bins, and an intrinsic shape dispersion of $\sigma_\epsilon = 0.3$ per component. For the correlation functions, noise only contributes via the covariance, which is described in Section 6.2.1.2 below.

6.2.1.2. Likelihood and covariance

A Gaussian likelihood was used for both the power spectra and correlation functions. This was shown to be sufficiently accurate for power spectra in Chapter 4. It can also be expected to be sufficiently accurate for correlation functions, since they are simply linear transformations of power spectra. To remove any unwanted randomness from the size of the posterior uncertainties, no observed realisation was simulated and instead the model data vector was used at the fiducial parameter values.

The covariance of the unbinned power spectrum was calculated assuming Gaussian fields, following Equation (4.6). This was then transformed linearly to obtain both the covariance of the binned power spectrum and the binned correlation function, using the general rule that for any data vector \mathbf{x} having covariance matrix Σ , the covariance of \mathbf{Mx} , where \mathbf{M} is the matrix representing some general linear transformation, is given by

$$\text{Cov}(\mathbf{Mx}) = \mathbf{M}\Sigma\mathbf{M}^\top. \quad (6.12)$$

For the power spectrum, noise was included in the unbinned C_ℓ s used to calculate the covariance, as described above in Section 6.2.1.1. For the correlation function, noise-free C_ℓ s were used, and a noise contribution to the diagonal of the covariance was added using the following expressions (Schneider et al. 2002; Joachimi, Schneider & Eifler 2008a; Heymans et al. 2013; Troxel et al. 2018)

$$\text{Cov}_{\text{Noise}}\left(\xi_{NN}^{ij}(\theta_1), \xi_{NN}^{kl}(\theta_2)\right) = \frac{1}{N_p^{ij}(\theta_1, \Delta\theta_1)} \delta_{\theta_1\theta_2} (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}); \quad (6.13)$$

$$\text{Cov}_{\text{Noise}}\left(\xi_+^{ij}(\theta_1), \xi_+^{kl}(\theta_2)\right) = \frac{\left(\sigma_\epsilon^i \sigma_\epsilon^j\right)^2}{N_p^{ij}(\theta_1, \Delta\theta_1)} \delta_{\theta_1\theta_2} (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}); \quad (6.14)$$

$$\text{Cov}_{\text{Noise}}\left(\xi_-^{ij}(\theta_1), \xi_-^{kl}(\theta_2)\right) = \frac{\left(\sigma_\epsilon^i \sigma_\epsilon^j\right)^2}{N_p^{ij}(\theta_1, \Delta\theta_1)} \delta_{\theta_1\theta_2} (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}); \quad (6.15)$$

$$\text{Cov}_{\text{Noise}}\left(\xi_{NE}^{ij}(\theta_1), \xi_{NE}^{kl}(\theta_2)\right) = \frac{\left(\sigma_\epsilon^i \sigma_\epsilon^j\right)^2}{N_p^{ij}(\theta_1, \Delta\theta_1)} \delta_{\theta_1\theta_2} \delta_{ik} \delta_{jl}. \quad (6.16)$$

In Equations (6.13)–(6.16), $i-l$ represent redshift bins, σ_ϵ^i and σ_ϵ^j are the intrinsic shape

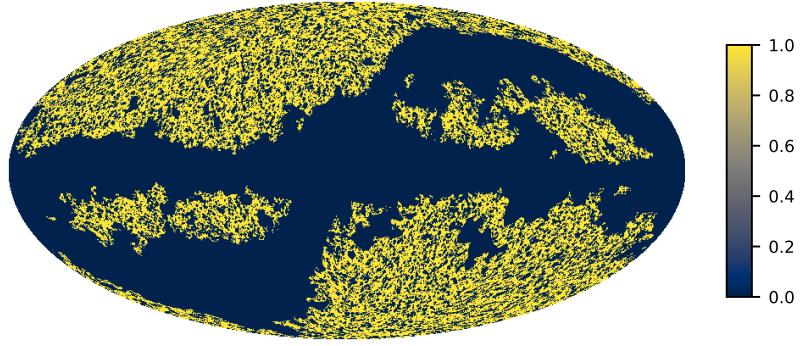


Figure 6.1. The Stage-IV-like mask describing a fictitious satellite-based survey used for the cut-sky results in this chapter.

dispersion per component in each redshift bin, and N_p is the number of galaxy pairs in a particular angular bin, given by (Friedrich et al. 2021)

$$N_p^{ij} = \frac{1}{2} A_{\text{survey}} A_{\text{bin}} N_i N_j, \quad (6.17)$$

where A_{survey} is the survey area in steradians, N_i and N_j are the galaxy number densities in bins i and j , and A_{bin} is the angular bin area given by

$$A_{\text{bin}} = 2\pi [\cos(\theta) - \cos(\theta + \Delta\theta)]. \quad (6.18)$$

6.2.1.3. Cut-sky treatment

For the cut-sky results, a Stage-IV-like mask describing a fictitious satellite-based survey was used. This was obtained by transforming the WMAP temperature mask used in Chapter 3 (Bennett et al. 2013), to excise the ecliptic plane in addition to the galactic plane. Additional point-source holes were added until a desired sky coverage of $f_{\text{sky}} = 0.3$ was achieved. The centre of each hole was chosen by selecting pixels at random. The immediate neighbouring pixels were also excised, after which a recursive probabilistic method was used, with a 50% chance of the next neighbouring pixels being removed, and so on up to a maximum hole radius of 6 pixels. The final mask is shown in Figure 6.1.

The impact of the sky cut on observed power spectra was modelled using a mixing matrix obtained using `NaMaster` (Alonso, Sanchez & Slosar 2019). A cut-sky power spectrum covariance was obtained using the improved narrow kernel approximation (Nicola et al. 2021) method described in Chapter 5. The sky cut does not affect the expected value of the correlation function, but does affect its covariance. This is modelled by applying a factor of $1/f_{\text{sky}}$ to the full-sky covariance matrix. Although this is presumably an approximation, it is standard practice for the cut-sky correlation function, used for example in the DES Year 3 covariance described in Friedrich et al. (2021). It was shown in Cabré et al. (2007) that this ‘ f_{sky} approximation’ is a good approximation to the simulated cut-sky correlation function

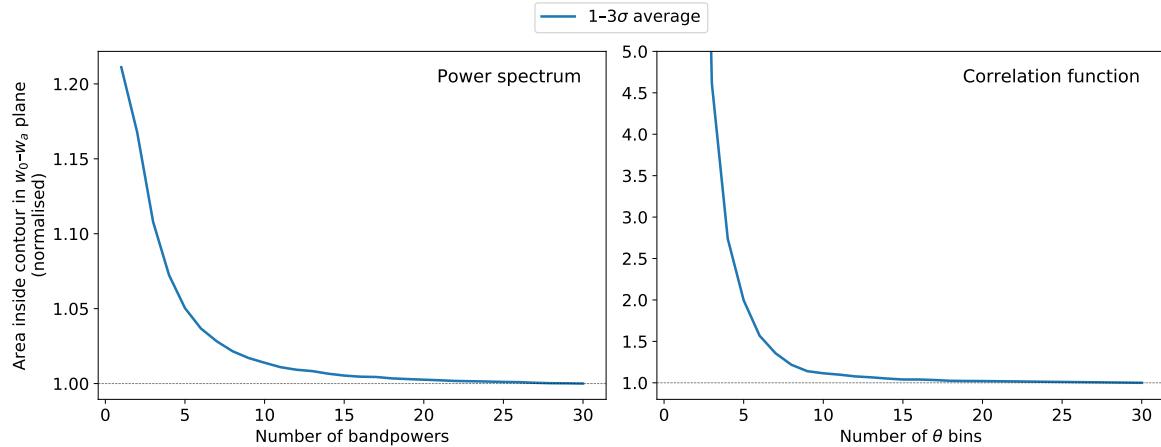


Figure 6.2. Dependence of posterior uncertainties in a joint full-sky analysis of w_0 and w_a on the number of angular bins for the power spectrum (left) and correlation function (right). Each line is averaged over 100 contours linearly spaced from 1σ to 3σ and normalised to be equal to 1 at its minimum, as described in the first paragraph of Section 6.2.2.

covariance, which is not the case for the power spectrum. Nevertheless, in Section 6.3.1 it is shown that the use of this approximation for the power spectrum does not significantly affect the dependence of posterior uncertainties on the number of angular bins. Therefore, it can be expected to be a sufficiently accurate approximation for the correlation function in this chapter.

6.2.2. Posterior uncertainty as a function of number of angular bins

The dependence of posterior uncertainties in a joint full-sky analysis of w_0 and w_a on the number of angular bins is shown for the power spectrum and correlation function in Figure 6.2. Each line is calculated by forming the joint posterior distribution of w_0 and w_a , drawing 100 credible regions linearly spaced from 1σ to 3σ (following the definitions of credible regions and sigma notation described in Section 2.4.1), measuring the area of every credible region as a function of the number of angular bins, normalising each curve to be equal to 1 at its minimum, and finally averaging over all 100 regions. This technique removes most of the noise that results from the finite resolution of the posterior grids.

Both lines in Figure 6.2 are mostly flat above around 10 angular bins, and converge to a minimum by around 20–25 bins. Below 10 bins, there is a sharp increase in contour area. For the power spectrum, however, this only reaches a maximum of around 20 per cent higher than the minimum area, even for a single bandpower. In contrast, the contour area for the correlation function diverges sharply at around 5 bins. This difference in behaviour is explored in Section 6.3.2.

To further illustrate the dependence shown in Figure 6.2, posterior distributions with different numbers of angular bins are shown in Figure 6.3 for the full-sky power spectrum

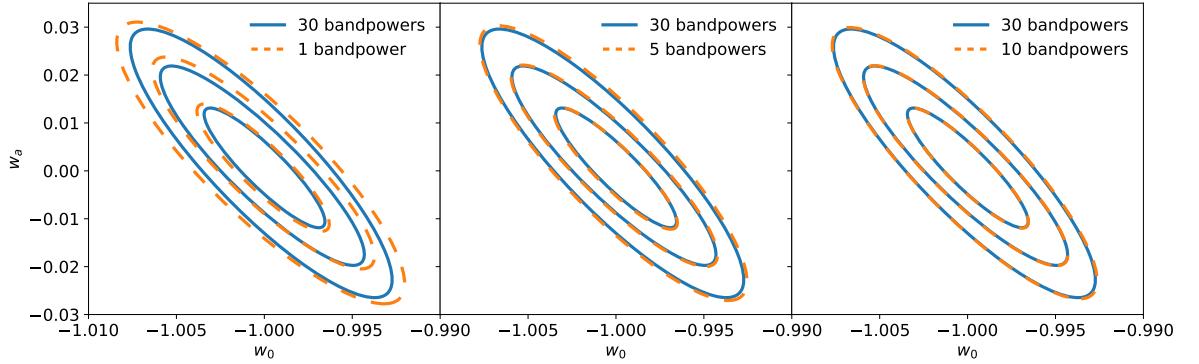


Figure 6.3. 1– 3σ posterior contours from the full-sky power spectrum with different numbers of bandpowers used in the likelihood analysis.

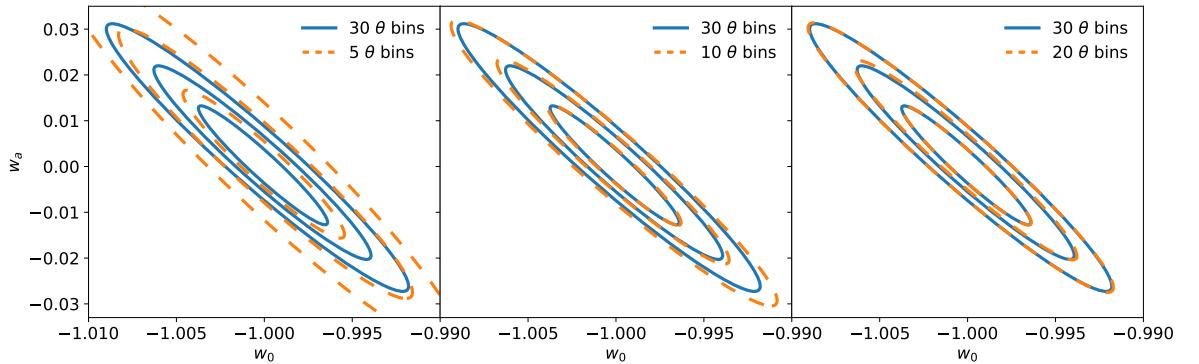


Figure 6.4. As Figure 6.3 but for the correlation function. 1– 3σ posterior contours from the full-sky correlation function with different numbers of angular bins used in the likelihood analysis.

and in Figure 6.4 for the full-sky correlation function. It can be seen in the left panel of Figure 6.3 that the single-bandpower posterior distribution is only slightly larger than the equivalent result with 30 bandpowers. The correlation function posteriors in Figure 6.4 show more degradation as the number of bins is decreased, as mentioned above.

Figure 6.5 shows the same dependence on the number of angular bins, but this time for a cut-sky analysis. The results are very similar to the full-sky results in Figure 6.2. For this reason, the remainder of this chapter considers only a full-sky setup, but the conclusions should be applicable to a cut-sky analysis as well.

Additional cosmological parameters are investigated in Figure 6.6. These are chosen to be the parameters included in the *Euclid* forecast paper (Euclid Collaboration: Blanchard et al. 2020): w_0 , w_a , Ω_m , Ω_b , σ_8 , n_s , h . These parameters are described in Chapter 1. Each parameter is shown for three different scale cuts. For the power spectrum, the maximum multipole ℓ_{\max} is varied from the baseline value of $\ell_{\max} = 2000$ to a more conservative $\ell_{\max} = 1000$ and a more optimistic $\ell_{\max} = 5000$. For the correlation function, the respective equivalent minimum separation θ_{\min} values are $\theta_{\min} = 0.1 \text{ deg}$, $\theta_{\min} = 0.2 \text{ deg}$, and $\theta_{\min} = 0.03 \text{ deg}$. While the different parameters behave broadly similarly, there are some differences. For

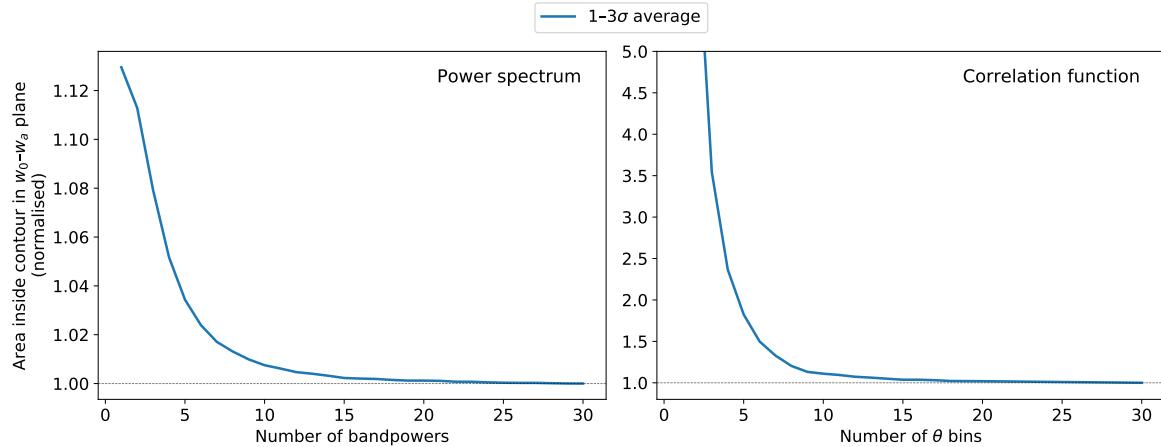


Figure 6.5. As Figure 6.2 but for a cut-sky analysis. Dependence of posterior uncertainties in a joint cut-sky analysis of w_0 and w_a on the number of angular bins for the power spectrum (left) and correlation function (right). Each line is averaged over 100 contours linearly spaced from 1σ to 3σ and normalised to be equal to 1 at its minimum, as described in the first paragraph of Section 6.2.2.

example, while some parameters such as w_0 only exhibit a 10–20 per cent degradation at low numbers of bandpowers, others such as Ω_b exhibit as much as 300 per cent degradation, perhaps because constraining such parameters requires sufficient angular resolution to resolve the shape of the power spectrum and not just its amplitude.

Ideally it would be possible to unify the results with different scale cuts in Figure 6.6 into a dependence on a single measure, such as the logarithmic angular bin size. However, the different parameters behave differently as ℓ_{\max} and θ_{\min} are varied, so this is not possible in practice. However, a certain subset of the parameters do exhibit such a convenient property for the power spectrum: it is shown in Figure 6.7 that for four parameters— w_0 , w_a , Ω_m and σ_8 —the number of bandpowers required for a given ℓ_{\max} value scales approximately as $\sqrt{\ell_{\max}}$. The reason for this particular dependence is not clear, but it is perhaps not a coincidence that these parameters are those which depend most on the amplitude of weak lensing power spectra and least on their shape. For this reason, these parameters are also the most strongly correlated in a *Euclid*-like analysis (Euclid Collaboration: Blanchard et al. 2020).

6.3. Exploration of additional effects

In this section, three different effects which may affect the required number of angular bins are explored. Section 6.3.1 investigates the impact of the f_{sky} approximation in the power spectrum covariance. Section 6.3.2 explores why the power spectrum and correlation function exhibit such different behaviour for low numbers of angular bins, while Section 6.3.3 studies the impact of noise.

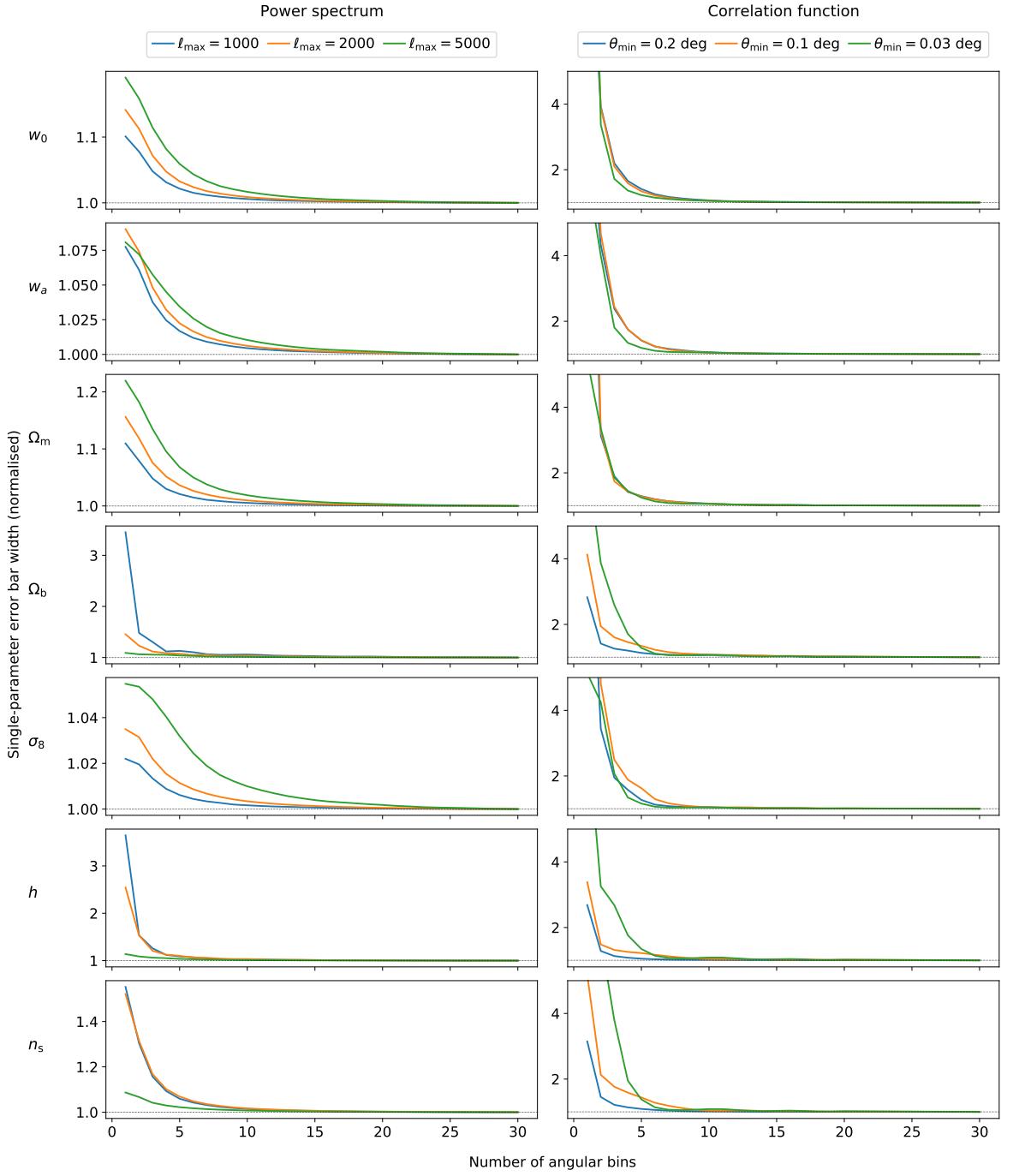


Figure 6.6. As Figure 6.2 but for a single parameter at a time. Dependence of posterior uncertainties in a single-parameter analysis on the number of angular bins for the power spectrum (left) and correlation function (right). Each line is averaged over 100 contours linearly spaced from 1σ to 3σ and normalised to be equal to 1 at its minimum, as described in the first paragraph of Section 6.2.2. Each parameter is constrained independently, with all other parameters held fixed. The results for different scale cuts are shown in different colours.

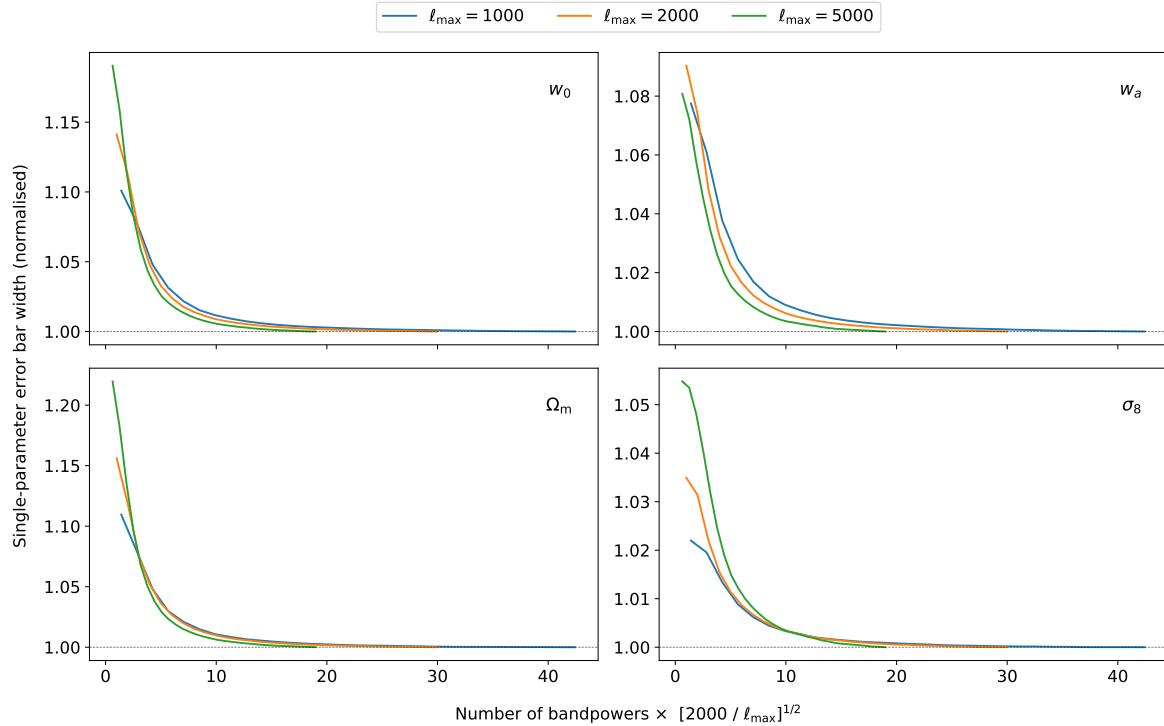


Figure 6.7. Dependence on the number of bandpowers for the full-sky power spectrum, for four parameters exhibiting a particular property: the number of bandpowers required for a particular level of degradation for different ℓ_{max} values approximately scales as $\sqrt{\ell_{\text{max}}}$.

6.3.1. Impact of f_{sky} approximation

In Section 6.2, the cut-sky power spectrum covariance was calculated using the Improved NKA method. A faster alternative is to use the ‘ f_{sky} approximation’:

$$\text{Cov}_{\text{cut-sky}} \approx \frac{1}{f_{\text{sky}}} \text{Cov}_{\text{full-sky}}, \quad (6.19)$$

where f_{sky} is the observed fraction of the sky. Figure 6.8 shows the dependence of posterior uncertainties in a joint cut-sky analysis of w_0 and w_a on the number of bandpowers when using the f_{sky} approximation compared to the more accurate Improved NKA method. While the curves produced by the two methods are not identical, the results are similar other than for a very coarse binning (1–2 angular bins). The good performance of the f_{sky} approximation for the power spectrum supports its use for the correlation function in Section 6.2, especially when combined with the results of Cabré et al. (2007), which show through a comparison with simulations that this approximation is much more accurate for the correlation function than it is for the power spectrum.

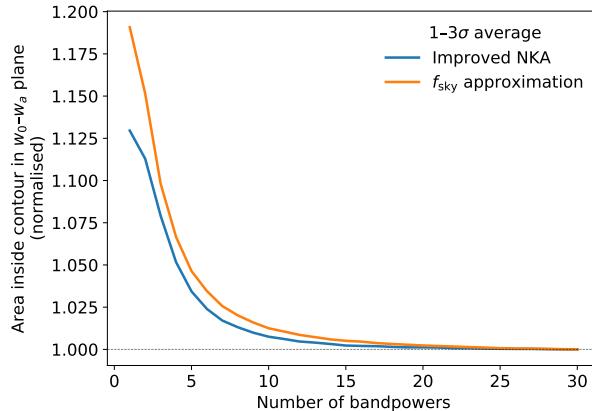


Figure 6.8. Dependence of posterior uncertainties in a joint cut-sky analysis of w_0 and w_a on the number of power spectrum bandpowers, when the covariance is calculated using the f_{sky} approximation (Equation 6.19) compared to the Improved NKA method. Each line is averaged over 100 contours linearly spaced from 1σ to 3σ and normalised to be equal to 1 at its minimum, as described in the first paragraph of Section 6.2.2.

6.3.2. Contrast between power spectrum and correlation function behaviour for small numbers of bins

A peculiar result that occurs throughout Section 6.2 is that posterior uncertainties obtained through a power spectrum analysis tend to degrade much less for small numbers of angular bins than those obtained through a correlation function analysis. For example, in Figure 6.2, even a single bandpower only returns a joint w_0 - w_a uncertainty around 20 per cent larger than with 30 bandpowers, whereas for the correlation function there is a sharp divergence by at least 500 per cent at around 5 angular bins. The explanation for this behaviour is the differing ways in which scales are weighted within each angular bin for the two methods, which will now be demonstrated.

The top panel of Figure 6.9 shows 45 theoretical power spectra, corresponding to the shear auto-power in the lowest redshift bin for different points in the w_0 - w_a plane. The points are chosen from a diagonal line running perpendicular to the w_0 - w_a degeneracy direction; that is, from the lower left to upper right of a given panel of Figure 6.3. Each model is coloured by its distance in σ from the fiducial model, determined from a full-sky unbinned power spectrum likelihood analysis. The fiducial model is $(w_0, w_a) = (-1, 0)$. In practice, the 45 power spectra all appear on top of one another. Error bars are shown at a selection of multipoles, and include a shape noise contribution. In the lower panel, the difference between each model and the fiducial model is shown, scaled by the error bar for each ℓ . This is effectively equivalent to the per- ℓ signal-to-noise as a function of scale. There is a region of very low signal-to-noise at low ℓ , $\ell \lesssim 50$, which rapidly increases to a peak at around $\ell \sim 200$, followed by a gentle decrease towards higher ℓ as shape noise begins to dominate.

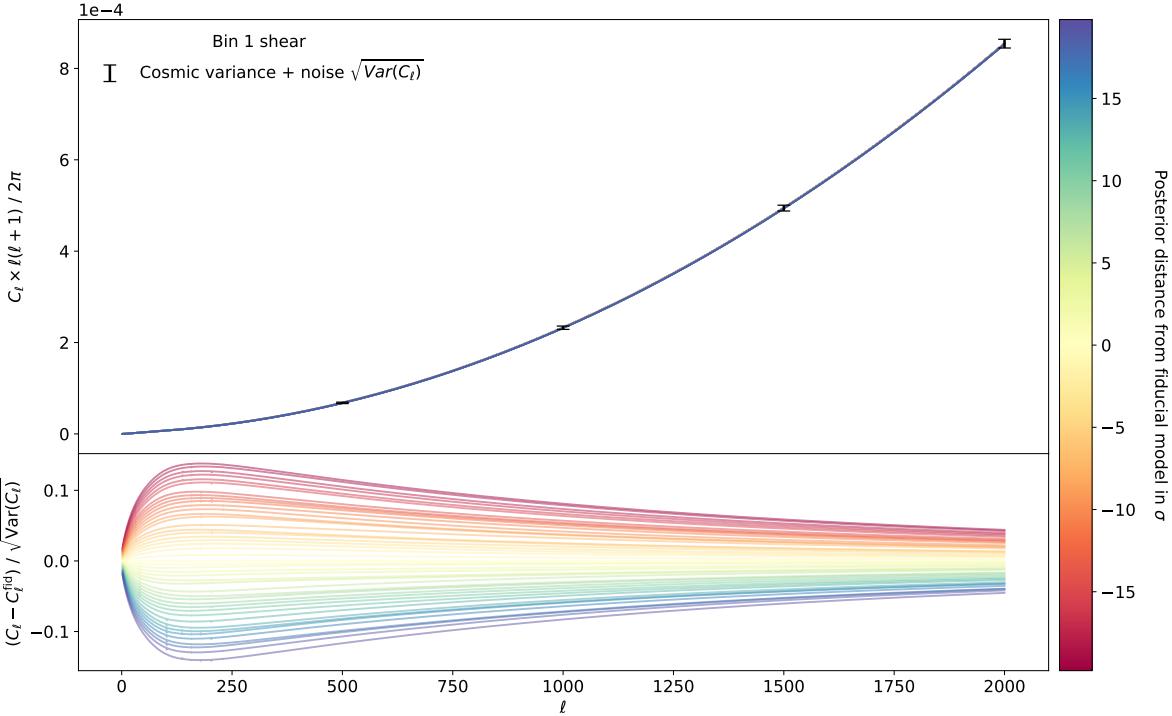


Figure 6.9. Top: 45 theoretical power spectra, corresponding to the shear auto-power in the lowest redshift bin for different points in the w_0-w_a plane. The points are chosen from a diagonal line running perpendicular to the w_0-w_a degeneracy direction. Each model is coloured by its distance in σ from the fiducial model, determined from a full-sky unbinned power spectrum likelihood analysis. In practice, the 45 power spectra all appear on top of one another. Error bars are shown at a selection of multipoles, and include a shape noise contribution. Bottom: The difference between each model and the fiducial model, scaled by the error bar for each ℓ .

The equivalent for the correlation function, for ξ^+ in the lowest redshift bin, is shown in Figure 6.10. 30 angular bins are used. The colour for each model is the same as in Figure 6.9. The signal-to-noise is roughly constant between 0.1 and 1 deg, before steadily decreasing to reach a very low level at 10 deg.

As seen in Figures 6.9 and 6.10, both the power spectrum and correlation function have poor signal-to-noise on large angular scales. This fact alone cannot explain the difference between them when small numbers of angular bins are used. The difference arises because of how scales are weighted within each bin. For the power spectrum, different ℓ within each bandpower are weighted as $\ell(\ell+1)$ following Equation (5.11), which quadratically downweights the largest scales, which are the lowest signal-to-noise part of the data vector. For the correlation function, however, different θ within each bin are weighted as $\sin \theta$, corresponding to a uniform distribution of galaxies (Friedrich et al. 2021). This weighting upweights the largest scales, which can dramatically decrease the signal-to-noise when combining scales into few bins.

The effect of the $\sin \theta$ weighting is small when relatively large numbers of angular bins are

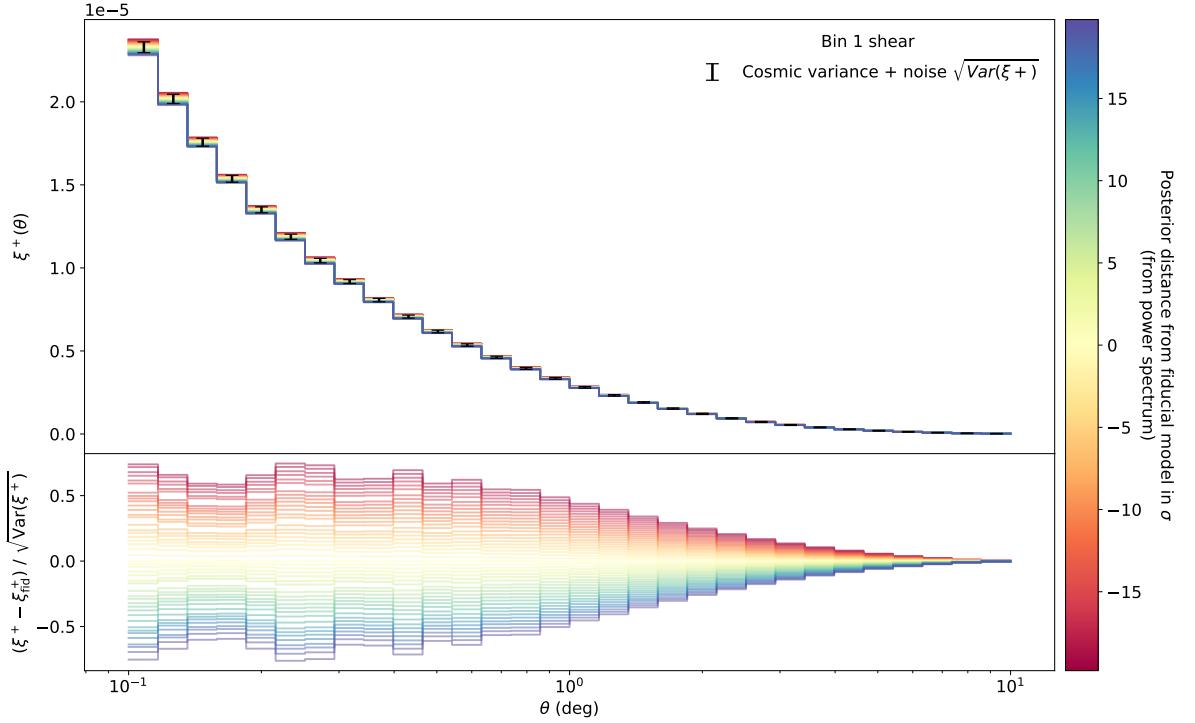


Figure 6.10. As Figure 6.9 but for the correlation function. Top: 45 theoretical correlation functions, corresponding to ξ^+ in the lowest redshift bin for different points in the w_0-w_a plane. The points are chosen from a diagonal line running perpendicular to the w_0-w_a degeneracy direction. Each model is coloured by its distance in σ from the fiducial model, determined from a full-sky unbinned power spectrum likelihood analysis. Error bars are shown for each angular bin, and include a shape noise contribution. Bottom: The difference between each model and the fiducial model, scaled by the error bar for each angular bin.

used. The left panel of Figure 6.11 shows the same selection of correlation functions and per-bin signal-to-noise for 30 angular bins, while the right panel uses 15 bins. Each angular bin on the right panel corresponds to exactly two bins on the left panel. It can be seen that the $\sin \theta$ weighting has little effect, because for instance the value of ξ^+ in the first bin in the right panel is approximately equal to the mean of the values of ξ^+ in the first two bins in the left panel. This is because with a fine binning such as that used here, the difference in $\sin \theta$ between adjacent bins is not significant.

However, with a very low number of bins the effect of the $\sin \theta$ weighting is large. Figure 6.12 shows the difference between using two angular bins in the left panel and just a single bin in the right panel. When combining the two bins into one, the value of ξ^+ is heavily weighted towards the larger-scale bin. The result is that the signal-to-noise drops by an order of magnitude, as is evident in the lower panels.

It is possible to demonstrate that the power spectrum would experience the same problem if large scales were not strongly downweighted. To do so, it is necessary to introduce a new technique, in order to combine the signal-to-noise across all angular bins while taking correlations between bins into account. This involves defining the covariance-weighted

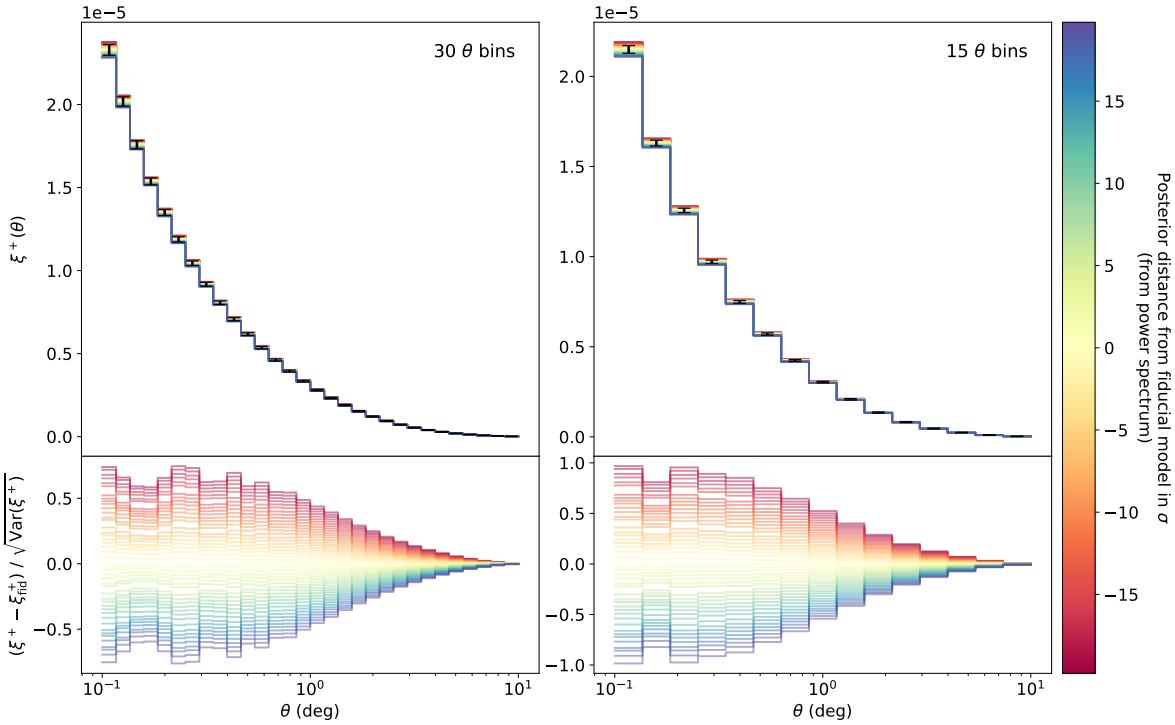


Figure 6.11. As Figure 6.10, but for 30 angular bins (left panel) compared to 15 (right panel).

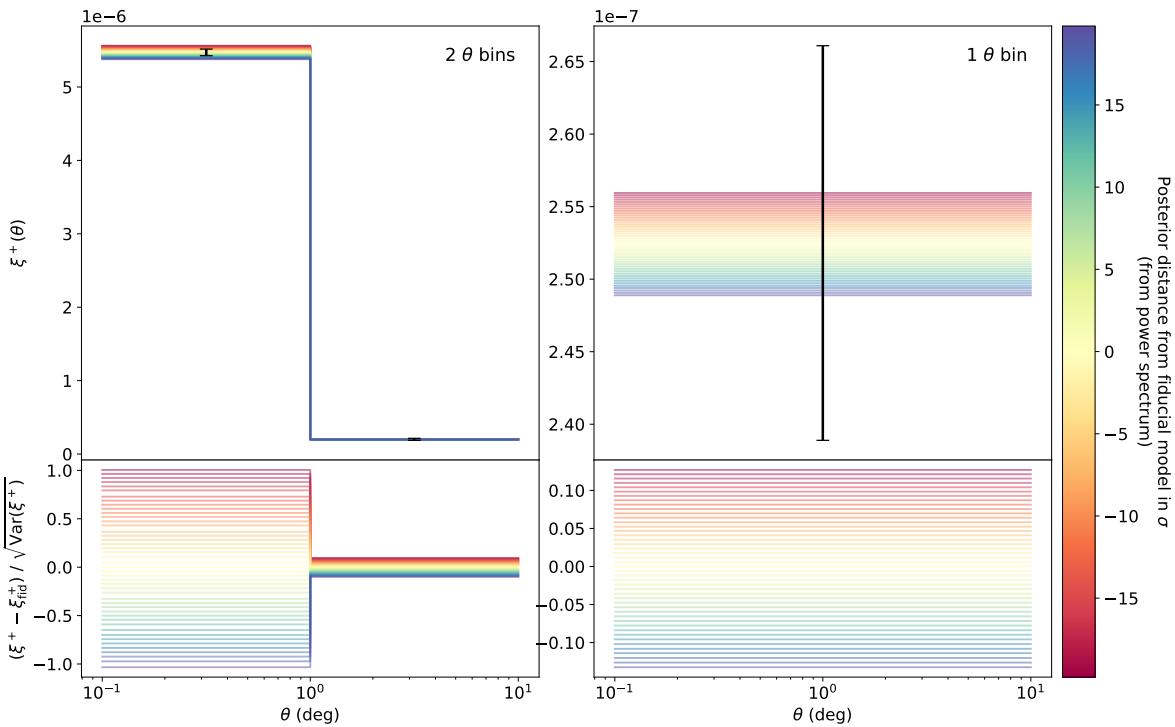


Figure 6.12. As Figure 6.10, but for 2 angular bins (left panel) compared to 1 (right panel).

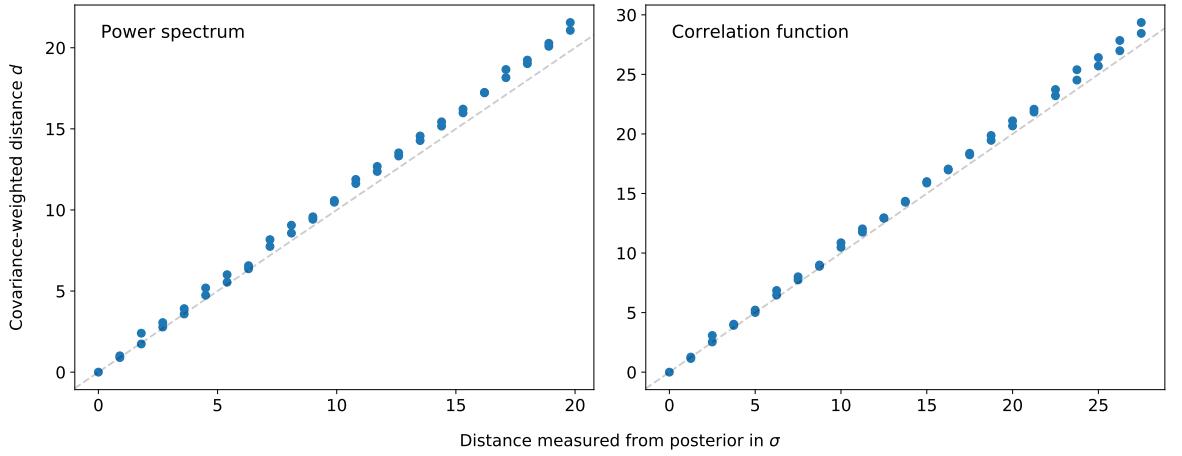


Figure 6.13. Comparison of the covariance-weighted distance d defined in (6.20) to the distance from the fiducial model measured from the posterior distribution resulting from a full likelihood analysis, for each of 45 different models. Each model corresponds to a different point in the w_0 - w_a plane, selected from a diagonal line running perpendicular to the w_0 - w_a degeneracy direction. The dashed line represents equality.

distance d^2 ² as

$$d^2 = (\mathbf{x} - \mathbf{x}^{\text{fid}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{x}^{\text{fid}}), \quad (6.20)$$

where \mathbf{x} is the theoretical data vector for a particular model (a set of C_ℓ or $\xi(\theta)$ values), \mathbf{x}^{fid} is the data vector predicted by the fiducial model, and $\boldsymbol{\Sigma}$ is the covariance matrix.

Figure 6.13 shows that Equation (6.20) correctly predicts the distance from the fiducial model measured from the posterior distribution resulting from a full likelihood analysis. The left panel shows the value of the covariance-weighted distance d against the distance from the fiducial model in σ measured from the posterior for the power spectrum, while the right panel shows it for the correlation function, for the selection of 45 models used throughout this section. The agreement is worse farther from the fiducial model, which may be a result of being close to the boundary of the prior region used in the likelihood analysis.

The upper left panel of Figure 6.14 shows the covariance-weighted distance from the fiducial model as a function of the number of angular bins for the power spectrum with the usual $\ell(\ell+1)$ weighting, for each of the 45 models. The results are consistent with those found from a full likelihood analysis in Figure 6.2: the distance at which each model is excluded only decreases slightly when reducing the number of bandpowers, even to a single bandpower. The upper right panel shows the equivalent for the correlation function, which is weighted within each bin as $\sin \theta$. As seen previously in Figure 6.2, there is a much more severe degradation in performance for small numbers of bins, with the distance from the fiducial model shrinking to approximately zero for a model excluded at more than 30σ with 20–30 angular bins.

² This distance d is equivalent to the Mahalanobis distance (Mahalanobis 1936).

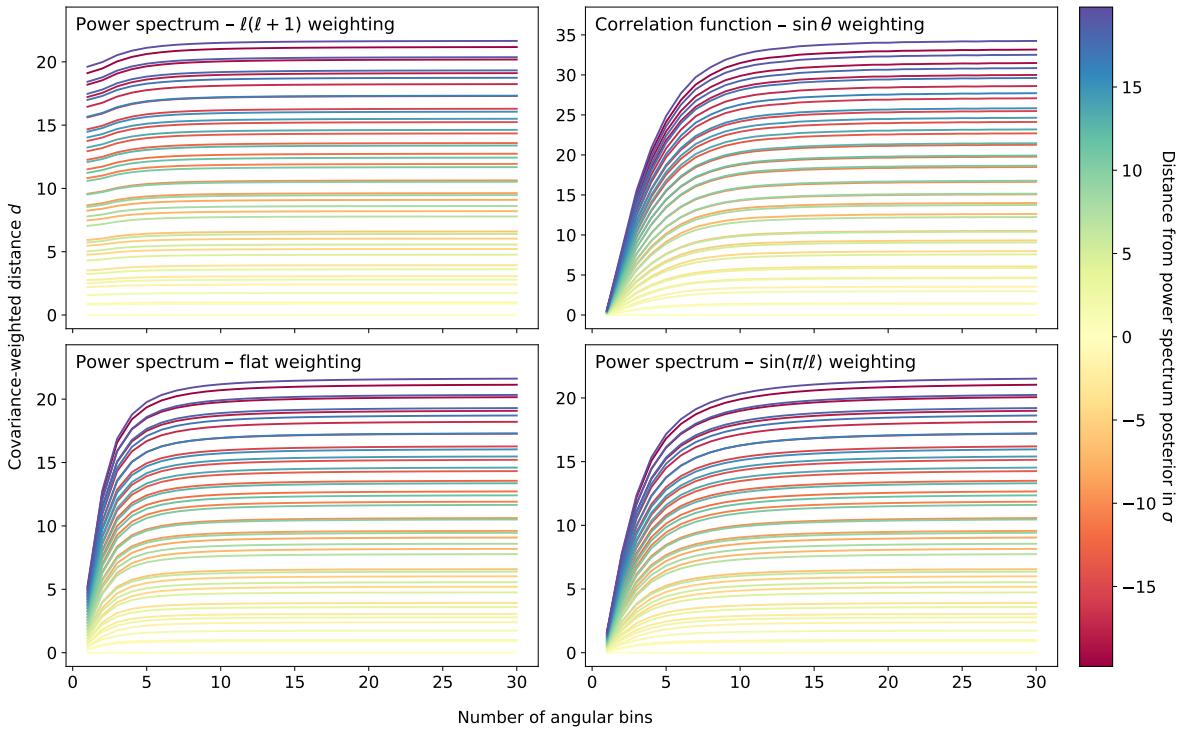


Figure 6.14. Covariance-weighted distance d from the fiducial model, as defined in Equation (6.20), as a function of the number of angular bins, for a selection of 45 different models. Each model corresponds to a different point in the w_0-w_a plane, selected from a diagonal line running perpendicular to the w_0-w_a degeneracy direction. Each panel explores a different weighting of scales within each angular bin. Top left: power spectrum with the usual $\ell(\ell+1)$ weighting. Top right: correlation function with the $\sin \theta$ weighting corresponding to a uniform distribution of galaxies. Bottom left: power spectrum with a flat weighting, where each ℓ is weighted equally within each bandpower. Bottom right: power spectrum with a correlation-function-like $\sin(\pi/\ell)$ weighting.

The lower left panel of Figure 6.14 shows the effect of applying a flat weighting to the power spectrum; that is, weighting each ℓ equally within a given bandpower. The effect here is to reduce the amount by which large scales are downweighted, which leads to a degradation in the overall signal-to-noise for small numbers of bandpowers. Finally, in the lower right panel, a weighting is applied that is equivalent to the way in which the correlation function is weighted: each ℓ is weighted as $\sin(\pi/\ell)$. The resulting form of the distance from the fiducial model as a function of the number of bins is almost identical to that for the correlation function. This demonstrates that the weighting of scales within bins is responsible for the different behaviour of the power spectrum and the correlation function for low numbers of angular bins.

6.3.3. Impact of noise

It is possible that the results derived in the main analysis presented in Section 6.2 may vary if the level of noise present in the data changes. This is investigated in this section.

The top panel of Figure 6.15 shows the posterior uncertainty in w_0 from a full-sky power spectrum analysis, as a function of the number of bandpowers, for three noise levels. The baseline noise level corresponds to 30 galaxies / arcmin² across all redshift bins, while the line labelled ‘ $\times 100$ noise’ corresponds to 0.3 / arcmin² and the line labelled ‘ $\times 0.01$ noise’ corresponds to 3000 / arcmin². Different behaviour is observed for the three different noise levels, which is discussed further below. Varying the noise by a factor 100 turned out to not be possible for the correlation function, due to problems with the conditioning of the covariance matrix for low noise levels. For a fair comparison with the correlation function, the lower left panel therefore shows the results for power spectrum with the noise varying by a factor of 2. ‘ $\times 2$ noise’ corresponds to 15 galaxies / arcmin², while ‘ $\times 0.5$ noise’ corresponds to 60 / arcmin². While the spacing between the three lines is smaller than in the top panel, it is still clearly present in the same arrangement.

The right panel of Figure 6.15 shows the equivalent figure for the correlation function, with a factor of 2 between each noise level. In this case, little clear difference is observed between the different noise levels.

Insight into the different behaviour of the power spectrum and correlation function under varying noise levels seen in Figure 6.15 may be gained by studying the signal-to-noise as a function of scale for each method with an enhanced level of noise. This may be studied using some of the methods developed in Section 6.3.2 above. Since the methods used here do not require the use of a covariance matrix, a factor of 100× may be used without any problems, which allows a clearer illustration of the effect of noise.

The lower panel of Figure 6.16 shows the per- ℓ signal-to-noise of the shear auto-power spectrum in the lowest redshift bin for the 100× baseline noise level. (See Section 6.3.2

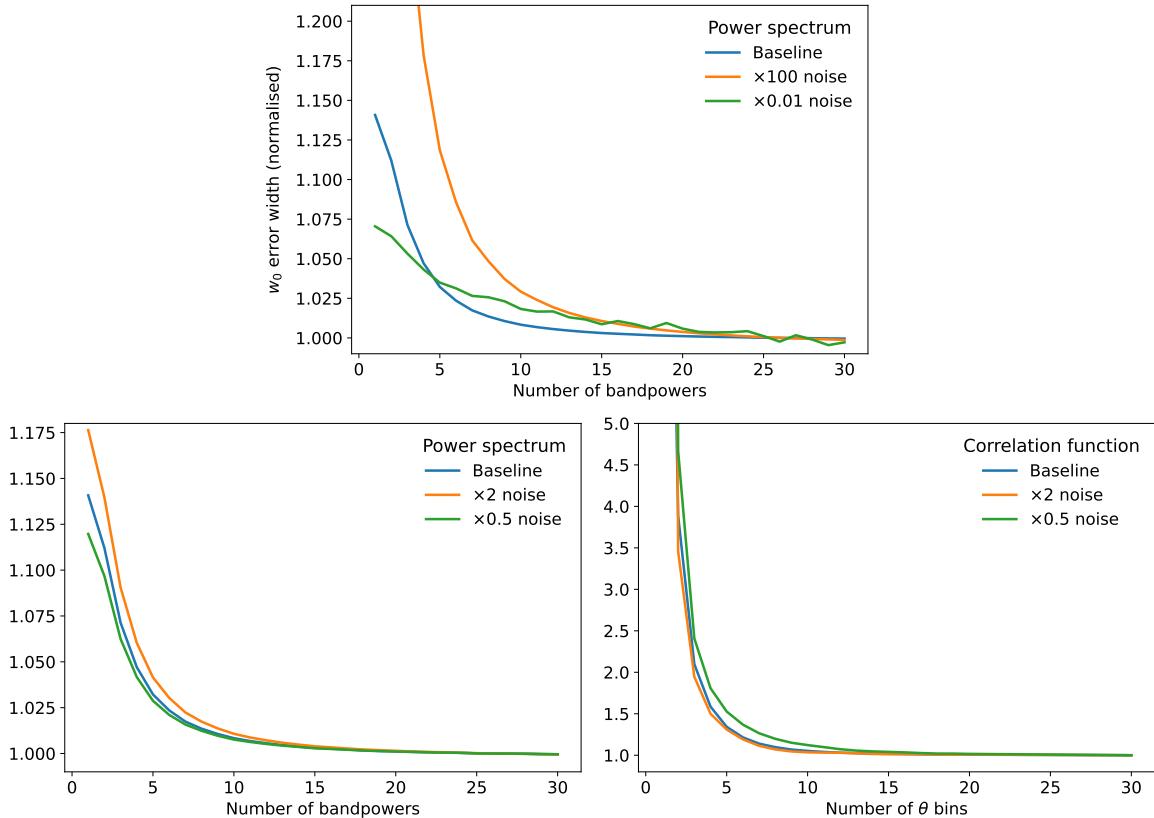


Figure 6.15. Dependence of w_0 posterior uncertainty on the number of angular bins for the power spectrum and correlation function, for three different noise levels. Top: power spectrum, varying noise by a factor of 100. Lower left: power spectrum, varying noise by a factor of 2. Lower right: correlation function, varying noise by a factor of 2. Each line is averaged over 100 contours linearly spaced from 1σ to 3σ and normalised to be equal to 1 at its minimum, as described in the first paragraph of Section 6.2.2. The baseline noise level corresponds to 30 galaxies / arcmin² across all redshift bins.

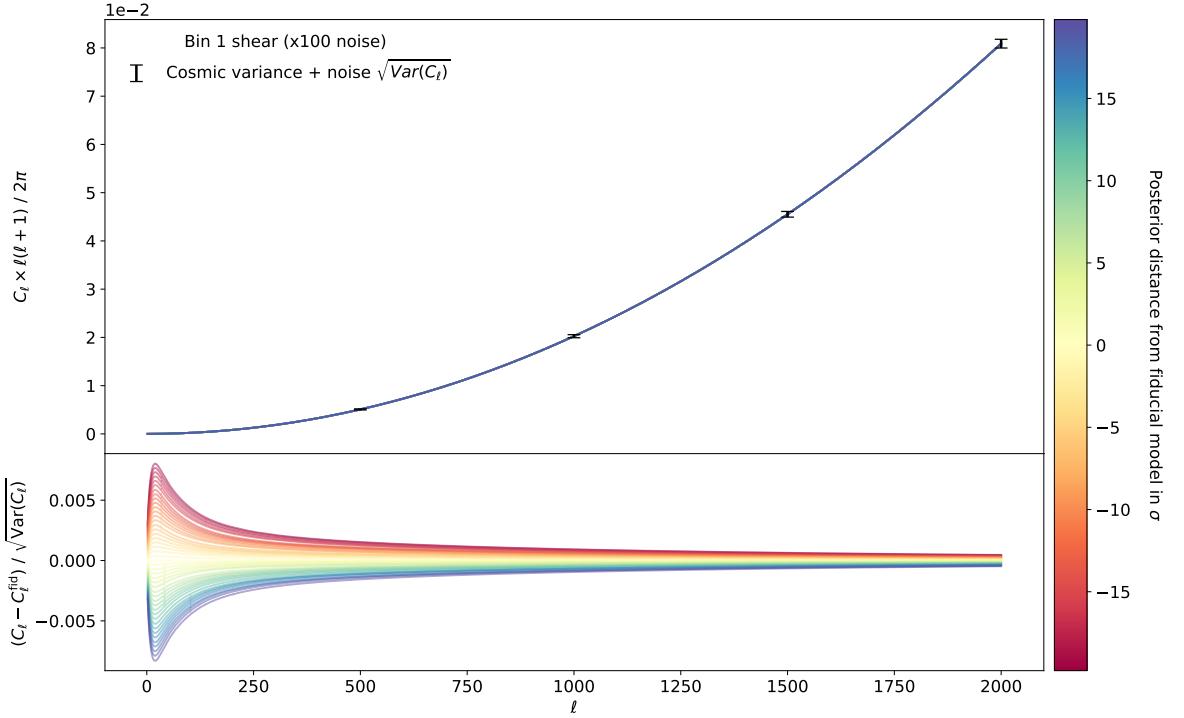


Figure 6.16. As Figure 6.9 but with $100\times$ baseline noise. Top: 45 theoretical power spectra, corresponding to the shear auto-power in the lowest redshift bin for different points in the w_0-w_a plane. The points are chosen from a diagonal line running perpendicular to the w_0-w_a degeneracy direction. Each model is coloured by its distance in σ from the fiducial model, determined from a full-sky unbinned power spectrum likelihood analysis. In practice, the 45 power spectra all appear on top of one another. Error bars are shown at a selection of multipoles, and include a shape noise contribution. Bottom: The difference between each model and the fiducial model, scaled by the error bar for each ℓ .

for a more detailed explanation of this type of figure.) In contrast to the relatively small drop in signal-to-noise towards higher ℓ seen in Figure 6.9, there is now a much larger decrease, in addition to a shift in the peak towards lower ℓ ($\ell \sim 50$). This is because the smaller scales are heavily suppressed by the increased noise level. If these low signal-to-noise small scales are combined in a single bin with the high signal-to-noise region, they will be strongly upweighted by the $\ell(\ell+1)$ weighting of ℓ s within bandpowers, leading to a strong decrease in the overall signal-to-noise. This is what is observed in the ' $\times 100$ noise' line in the top panel of Figure 6.15, and to a smaller extend in the ' $\times 2$ noise' line in the lower left panel of the same figure. The opposite effect can be expected to occur for low noise levels, which can explain the shallow slope of the ' $\times 0.01$ noise' and ' $\times 0.5$ noise' lines.

Figure 6.17 shows the equivalent for the correlation function. Compared to the baseline noise level seen previously in Figure 6.10, there is little change in shape when the noise is increased. This is unlike the power spectrum, and explains why the dependence on the number of correlation function bins sees little change when the noise level is varied.

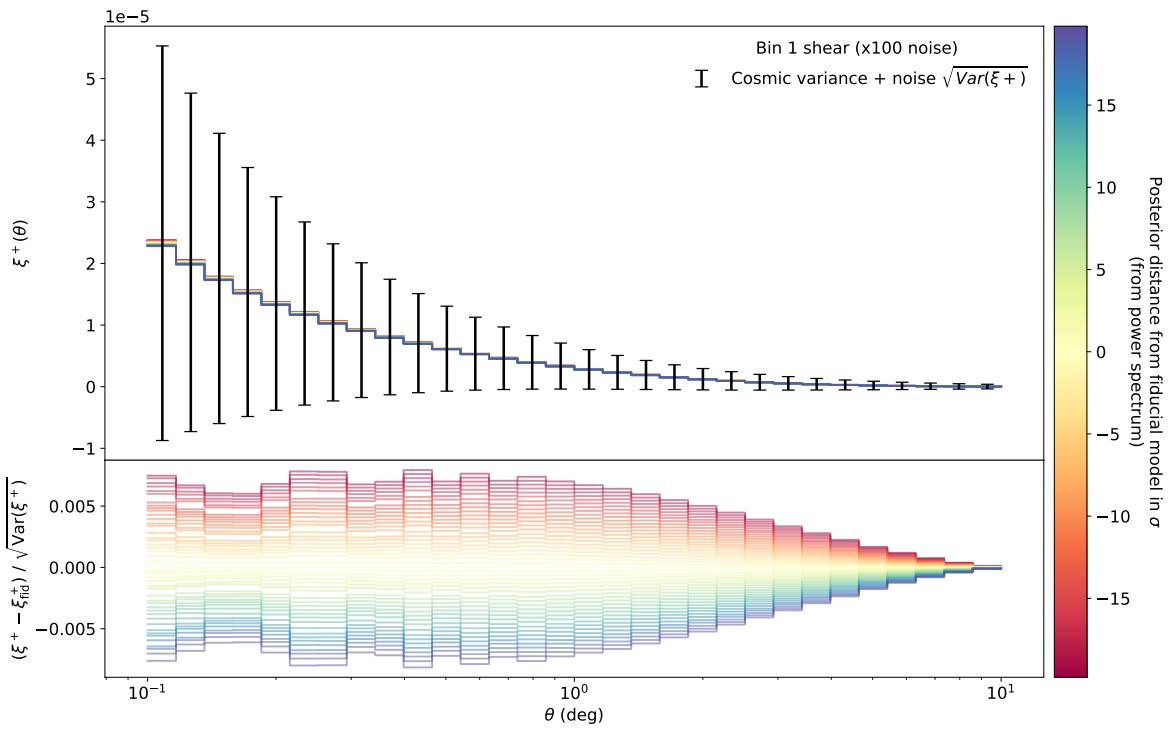


Figure 6.17. As Figure 6.10 but with $100\times$ baseline noise. Top: 45 theoretical correlation functions, corresponding to ξ^+ in the lowest redshift bin for different points in the w_0-w_a plane. The points are chosen from a diagonal line running perpendicular to the w_0-w_a degeneracy direction. Each model is coloured by its distance in σ from the fiducial model, determined from a full-sky unbinned power spectrum likelihood analysis. Error bars are shown for each angular bin, and include a shape noise contribution. Bottom: The difference between each model and the fiducial model, scaled by the error bar for each angular bin.

6.4. Conclusions

This chapter has investigated the question of how the statistical uncertainties in the posterior distributions of cosmological parameters derived from weak lensing two-point statistics depend on the number of angular bins used in the analysis. In Section 6.2 it was shown that for some parameters such as w_0 and w_a , these uncertainties for the power spectrum only increase by 7–20 per cent even for a single bandpower, relative to 30 bandpowers. In contrast, for the correlation function there is a sharp divergence by several hundred per cent at around 5 angular bins. However, these results vary quite substantially depending on the parameters in question and other aspects of the analysis. For example, power spectrum constraints in Ω_b and h can degrade by over 200 per cent for low numbers of bandpowers, while the same parameters do not suffer quite as badly for the correlation function. There is an inconsistent dependence on scale cut: a selection of four cosmological parameters ($w_0, w_a, \Omega_m, \sigma_8$) exhibit a dependence of $\sqrt{\ell_{\max}}$ for the power spectrum, but not for the correlation function, while the other parameters investigated (Ω_b, h, n_s) show no such dependence.

In Section 6.3 a number of additional effects have been investigated. It was shown in Section 6.3.1 that the use of the ‘ f_{sky} approximation’ for the power spectrum covariance leads to similar results to using the full Improved NKA method. The different behaviour of the power spectrum and correlation function for small numbers of bins was investigated in Section 6.3.2, where it was shown that this arises from the different ways in which scales are weighted within each angular bin. Finally, the effect of noise was studied in Section 6.3.3, where it was shown that the number of angular bins needed for the correlation function has no strong dependence on the level of noise, but that a higher noise level in the power spectrum requires more bandpowers.

While there is no single rule for the number of required angular bins that captures all possible analysis setups, the results in this chapter suggest that, for instance, if a maximum degradation on parameter uncertainties of 10 per cent is required, then 10 bandpowers appears comfortably sufficient for power spectra, and 15 angular bins for correlation functions. This is true for any angular range up to at least $\ell_{\max} = 5000$ ($\theta_{\min} = 0.03$ deg), although significantly higher scale cuts than this may require more bins.

Chapter 7

Weak lensing estimation with convolutional neural networks

7.1. Introduction

Weak lensing, as introduced in Chapter 1, is a probe with enormous potential to explore the low-redshift evolution of structure in the Universe and constrain theories of dark energy. However, it is accompanied by many observational and analytical challenges, and sources of bias and systematic error, some of which were described in Chapter 1. These challenges differ depending on the wavelength of the observations. Chapters 3–6 have been written in the context of an optical imaging survey, such as *Euclid*. Radio observations, on the other hand, bring their own challenges. One such challenge is reliable galaxy shape measurement. The fundamental difference between optical and radio data is that the raw data for radio observations are in the form of Fourier-space visibilities. One option for shape measurement from radio visibilities is to first produce images and then to use the same shape estimation techniques as used in optical weak lensing surveys. This approach was taken in Hillier et al. (2019), for instance. However, as pointed out in that paper, the imaging process may introduce systematic errors into the data, and this approach fails to fully realise the benefit of having a known point spread function (PSF), which is a key advantage of radio observations for weak lensing. As a result, methods have been developed for estimating galaxy shapes, or weak lensing shear, directly from radio visibilities (Rivi et al. 2016; Rivi & Miller 2018; Rivi et al. 2019; Hillier 2020; Harrison et al. 2020). However, these techniques are not yet known to be sufficiently mature and reliable for use with data from the upcoming Square Kilometre Array radio observatory (SKA; described in Chapter 1), which will achieve an unprecedented level of statistical precision due to the large number of galaxies predicted to be observed (Brown et al. 2015). The topic of shear estimation from radio visibilities therefore remains an active area of research.

One possible route to overcoming this challenge may be provided by machine learning. Machine learning, in the broadest sense, is a class of methods which involve describing some function using one or more parameters and then finding appropriate values for these parameters using a ‘training’ set of known pairs of input and output values for the function.¹ A simple example is a linear regression, in which the slope and intercept of the line are the two parameters to be ‘learned’. A more complex class of model (though still funda-

¹ In fact, the definition of machine learning is even broader than this, as it includes unsupervised methods in which a training set is not present. Unsupervised learning is not discussed further within this chapter.

mentally simple, as described in Section 7.2) are neural networks, which have recently exploded in popularity due to the unprecedented availability of training data and computing resources. Within astrophysics and cosmology, popular applications include radio galaxy classification (Aniyan & Thorat 2017; Scaife & Porter 2021; Mohan et al. 2022), supernova classification (Charnock & Moss 2017; Möller & de Boissière 2020), star–galaxy discrimination (Kim & Brunner 2017), neutron star nuclear mass prediction (Utama, Piekarewicz & Prosper 2016; Niu & Liang 2018), strong lensing analysis (Hezaveh, Levasseur & Marshall 2017; Petrillo et al. 2017; Perreault Levasseur, Hezaveh & Wechsler 2017; Jacobs et al. 2017), real-time gravitational wave analysis (George & Huerta 2018), photometric redshift estimation (Bonnett 2015; Hoyle 2016; Pasquet et al. 2019) and efficient likelihood sampling (Manrique-Yus & Sellentin 2020). The work presented in this chapter uses convolutional neural networks, a type of neural network that uses convolutions, which are introduced in Section 7.2.2.

As a simpler starting point from which to work towards the application of radio galaxy weak lensing, this chapter focuses on weak lensing of the cosmic microwave background (CMB; described in Chapter 1), starting with lensed CMB maps (i.e. neglecting the CMB map-making process, which is somewhat analogous to radio imaging in the challenges it brings). This is an exploratory project, which proceeds by starting with an extremely simplified problem and gradually adding complexity, with an aim to ultimately converge towards a realistic application. The project as presented here is, in this respect, a work in progress. Nevertheless, it contains some useful insights into what may be achieved using convolutional neural networks for weak lensing estimation. This can complement the existing literature on the topic: Ribli, Dobos & Csabai (2019) and Tewes et al. (2019) both demonstrated the potential of neural networks for galaxy shape and shear estimation for optical galaxy imaging surveys, while Nurbaeva et al. (2015) showed the application to PSF deconvolution, and Ribli et al. (2019) used neural networks for map-based parameter inference. This work has some overlap with Caldeira et al. (2019), who tested the use of convolutional neural networks to learn the functional mapping from maps of Stokes parameters describing CMB polarisation to lensing convergence maps, and found a superior performance compared to a quadratic estimator approach.

An overview of the theory of neural networks is presented in Section 7.2, focusing specifically on convolutional neural networks in Section 7.2.2. The methods used to generate training data and to build and train convolutional neural network models are described in Section 7.3. Section 7.4 describes the iterative exploration, starting from an extremely simplified problem and gradually adding additional complexity and realism, detailing the models used and the results obtained at each stage. A discussion of the status of the project and its achievements and challenges is presented in Section 7.5.

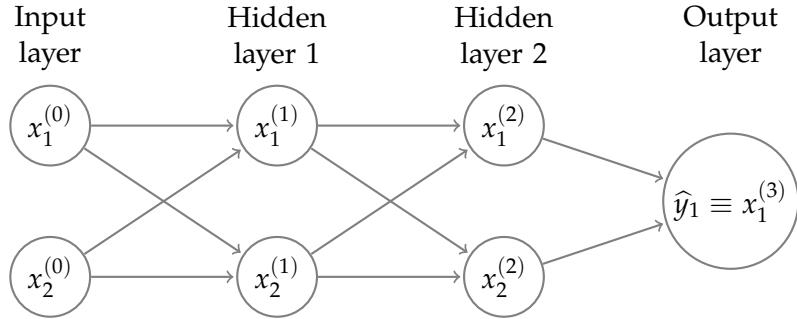


Figure 7.1. A simple neural network, containing four layers: the input layer, two hidden layers, and the output layer. Each node represents a single number, so this network represents a mapping from two numbers ($x_1^{(0)}, x_2^{(0)}$) to a third (y_1).

7.2. Theory

An overview of neural networks and how they are trained is given in Section 7.2.1, before convolutional neural networks are introduced in Section 7.2.2, along with some of the specific methods, such as loss functions and optimisation algorithms, that are used in this work.

7.2.1. Neural networks

A neural network, in the most basic sense, is a method to approximately emulate any function by the repeated alternation of simple linear and non-linear functions. It is formed as a series of layers, where each layer contains one or more nodes. A node constitutes a single number, for any given input passed through the network. A simple example is shown in Figure 7.1, showing one input layer with two nodes, two hidden layers, each with two nodes, and an output layer with a single node. This could represent a functional mapping from two numbers ($x_1^{(0)}, x_2^{(0)}$) to a third (y_1).

Mathematically, the nodes within each layer can be represented as a vector; for example, for layer j containing two nodes,

$$\mathbf{x}^{(j)} = \begin{pmatrix} x_1^{(j)} \\ x_2^{(j)} \end{pmatrix}. \quad (7.1)$$

The output from layer j is then given as the product of a matrix of weights $\mathbf{W}^{(j)}$ and the output from the previous layer $\mathbf{x}^{(j-1)}$, added to a further vector of ‘bias’ weights $\mathbf{b}^{(j)}$ and passed through a non-linear activation function f , as

$$\mathbf{x}^{(j)} = f \left(\mathbf{W}^{(j)} \mathbf{x}^{(j-1)} + \mathbf{b}^{(j)} \right). \quad (7.2)$$

Generally, the activation function f is selected by the user as a hyperparameter.² This leaves the linear weights $\mathbf{W}^{(j)}$ and $\mathbf{b}^{(j)}$, for all layers j , to be learned during training.

The number of layers, referred to as the ‘depth’ of the network, and the number of nodes in each layer—the ‘width’, which is generally different for each layer—are chosen by the user, depending on the nature of the function to be learned and the available resources.

7.2.1.1. Training process

The weights $\mathbf{W}^{(j)}$ and $\mathbf{b}^{(j)}$ are initialised randomly. Generally all initial elements are drawn from some joint distribution chosen by the user, as opposed to being initialised independently. This choice of initialisation is discussed in Section 7.2.2 below. The network is trained by being exposed to a (generally large) number of training samples, each containing an input $x^{(0)}$ and a known ‘truth’ output y . Each sample input in the training set is passed through the network to generate an estimate for y , denoted \hat{y} . This estimate is compared to the truth using a loss function $\mathcal{L}(y, \hat{y})$, which is chosen by the user—choices of loss function are discussed in Section 7.2.2 below.

The aim of the training process is to minimise the loss function. To achieve this, the derivative of the loss function with respect to each weight in the network is calculated, using a process called ‘back propagation’. This is simply the repeated application of the chain rule—for each layer j ,

$$\frac{d\mathcal{L}}{dx_i^{(j)}} = \sum_{i'} \frac{\partial x_{i'}^{(j+1)}}{\partial x_i^{(j)}} \frac{d\mathcal{L}}{dx_{i'}^{(j+1)}}. \quad (7.3)$$

Weights are then updated, generally in the direction of negative $d\mathcal{L}$ but with some amount of randomness to attempt to avoid outcomes such as becoming stuck in local minima or jumping over global minima. The exact behaviour depends on the choice of optimisation algorithm used, which is discussed in Section 7.2.2 below. Updating of weights occurs once per training ‘batch’, where a batch is a group of training samples of some predetermined size chosen by the user. A smaller batch size may allow faster learning, but is also more susceptible to noise caused by features of individual training samples that do not generalise to the whole training set.

A single pass through the entire training set is termed an ‘epoch’ of training. Generally the training process will continue for multiple epochs, until either some predetermined number of epochs or amount of training time is reached, or some convergence condition is fulfilled. This condition may involve the validation loss, \mathcal{L}_{val} , which is the loss function evaluated on a separate ‘validation’ set that is not used for training. This is typically evaluated at the end of each epoch, and can allow for the detection of overtraining, whereby the network is learning specific features of the training set that do not generalise to the underlying data. The results of the validation loss are not used to update the weights in

² A hyperparameter is a parameter that is selected by the user, rather than being learned during training.

the network, and therefore provide a more unbiased estimate of the true performance on the underlying data.

7.2.2. Convolutional neural networks

Convolutional neural networks (CNNs) are an extension to the simple case considered above, which are convenient when dealing with images in scenarios where the function to be learned is translationally invariant. In a CNN, the input to each layer is two-dimensional,³ denoted here for node i of layer j as $\mathbf{X}_i^{(j)}$, and the inner product in Equation (7.2) is replaced by a convolution with some kernel $\mathbf{K}_{ii'}^{(j)}$,

$$\mathbf{X}_i^{(j)} = f \left(\sum_{i'} \mathbf{K}_{ii'}^{(j)} * \mathbf{X}_{i'}^{(j-1)} + \mathbf{B}_i^{(j)} \right). \quad (7.4)$$

Each element of the bias vector $b^{(j)}$ from Equation (7.2) has now been replaced by a matrix $\mathbf{B}_i^{(j)}$, although sometimes this is reduced to a single value $b_i^{(j)}$, in which case

$$\mathbf{B}_i^{(j)} = b_i^{(j)} \mathbf{I}, \quad (7.5)$$

where \mathbf{I} is the identity matrix. Similarly, the activation function f is generally a non-linear matrix-to-matrix function, but is often chosen to be an elementwise scalar function. It must remain non-linear, however, since it is only by this property that the CNN as a whole may emulate a general non-linear function. As before, its choice is typically made by the user as a hyperparameter, and it is again the linear weights $\mathbf{K}_{ii'}^{(j)}$ and $\mathbf{B}_i^{(j)}$ for each (i, i', j) that are learned during training.

Each of the hyperparameter choices mentioned above will now be discussed in turn.

Kernel and bias initialisation

Every element of the convolution kernels $\mathbf{K}_{ii'}^{(j)}$ and bias matrices $\mathbf{B}_i^{(j)}$ must be initialised to some value at the start of training. The method used in this work was orthogonal kernel initialisation, whereby each kernel is initialised to be an orthogonal matrix, i.e. one satisfying

$$\mathbf{K}_{ii'}^{(j)} \left(\mathbf{K}_{ii'}^{(j)} \right)^T = \mathbf{I} \quad (7.6)$$

for all (i, i', j) . The elements are generated randomly under this constraint by first drawing each independently and identically from a Gaussian distribution and then taking a QR decomposition, following the method described in Saxe, McClelland & Ganguli (2013). Alternative kernel initialisation methods include Glorot uniform (explored briefly in Sec-

³ Higher dimensions are possible, but only the two-dimensional case is considered here. In many applications of CNNs the input data are multi-channel, but in most cases the channels are convolved independently and so each may be considered as a separate two-dimensional image.

tion 7.4.1 and described therein), and drawing elements independently from a Gaussian distribution with no subsequent transformation. The bias matrices were initialised to zero.

Activation function

The Rectified Linear Unit (ReLU) activation function was used in this work. This is an elementwise function that simply returns, for each element x ,

$$f_{\text{ReLU}}(x) = \begin{cases} x & \text{for } x > 0; \\ 0 & \text{otherwise.} \end{cases} \quad (7.7)$$

This minimally non-linear function is generally sufficient for the CNN as a whole to be able to emulate any non-linear function, given a sufficiently complex architecture and volume of training (e.g. Daubechies et al. 2019). One alternative choice is the sigmoid function,

$$f_{\text{sigmoid}}(x) = \frac{1}{1 + e^{-x}}, \quad (7.8)$$

but this has the shortcoming of becoming almost indistinguishable from unity as x approaches 1, which can significantly slow training.

Loss function

The loss function used in this work was the mean squared error,

$$\mathcal{L}_{\text{MSE}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \sum [\hat{\mathbf{Y}} - \mathbf{Y}]^2, \quad (7.9)$$

where the sum is over all elements. An alternative choice could be the mean absolute error,

$$\mathcal{L}_{\text{MSE}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \sum |\hat{\mathbf{Y}} - \mathbf{Y}|, \quad (7.10)$$

which might be suitable if a reduction in the sensitivity to outliers was required.

Optimisation algorithm

The optimisation algorithm determines how the weights in the network should be updated, given the result of the loss function and its derivatives for a given batch of training data. The ‘Adam’ optimiser (derived from ‘adaptive moment estimation’) was used in this work. This is a type of stochastic gradient descent method, in that it tends towards minimising the loss function, but with an element of stochasticity that helps to balance speed of convergence with risk of misconvergence. This stochasticity is not generated artificially, but instead arises naturally from the random noise present in small batches of training data, relative to the ‘signal’ of the underlying features that the network is intended to learn. The Adam optimiser keeps track of the following vectors of weights (elements of the convolu-

tional kernels and bias matrices) and their gradients and moments (all at step t):

- Vector of weights θ_t ;
- Vector of gradients g_t ;
- Vector of biased first moment estimates m_t ;
- Vector of biased second raw moment estimates v_t .

These are each updated at each step as follows (Kingma & Ba 2014):

$$g_t = \nabla_g \mathcal{L}(\theta_{t-1}); \quad (7.11)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t; \quad (7.12)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (g_t)^2; \quad (7.13)$$

$$\theta_t = \theta_{t-1} - \alpha_t \frac{m_t}{\sqrt{v_t + \epsilon}}, \quad (7.14)$$

where the training rate α , exponential decay rates β_1 and β_2 , and numerical stability constant ϵ are set by the user (see Section 7.4.1 for the values used in this work), and α_t is given by

$$\alpha_t = \alpha \frac{\sqrt{1 - (\beta_2)^t}}{1 - (\beta_1)^t} \quad (7.15)$$

(with superscripts here and above denoting powers, not labels).

CNNs are commonly used for image recognition and other classification-based computer vision applications (e.g. Dieleman, Willett & Dambre 2015; Chen et al. 2016; Howard et al. 2017), or for image-to-number regression problems (e.g. Nibali et al. 2018; Baldi et al. 2019; Pyo et al. 2019). However, in this work they are used for image-to-image regression: to emulate a function that takes an image (a lensed CMB map) as input and produces another image (the corresponding convergence map) as output. This motivates some of the choices described above, such as the use of a mean squared error loss function. The methods used to generate training data and to build and train CNN models are described below, in Section 7.3.

7.3. Method

7.3.1. Data generation

Training, validation and test data were generated using a simulated implementation of the basic weak gravitational lensing equation,

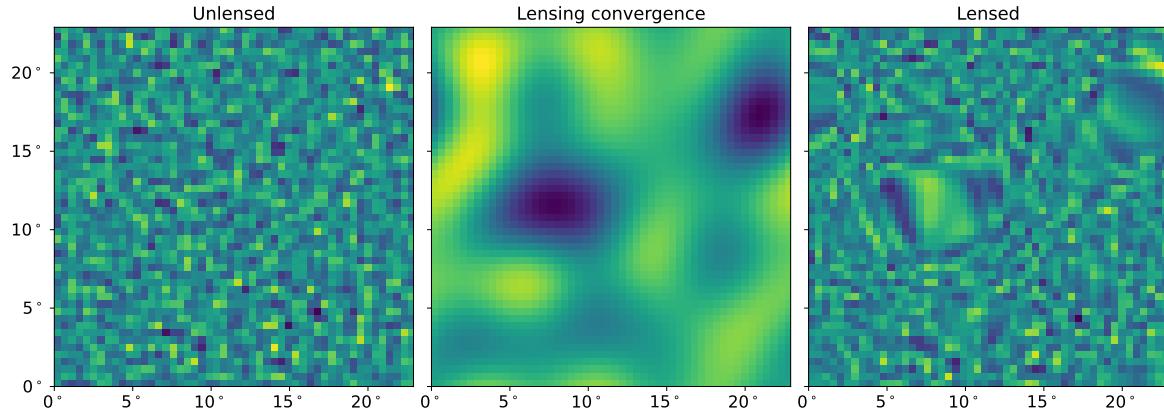


Figure 7.2. Demonstration of the CMB lensing process, where the convergence is exaggerated by a factor 30. The left panel shows the unlensed CMB field, the centre panel shows the convergence field (limited to $\ell_{\text{max}} = 50$), and the right panel shows the lensed CMB field. Each panel covers a field of view of 22.9 deg along each side.

$$T_{\text{lensed}}(\Omega) = T_{\text{unlensed}}(\Omega + \nabla\phi), \quad (7.16)$$

where $\phi(\Omega)$ is the lensing potential field defined in Equation (2.34). In this work, the lensing fields were defined not from $\phi(\Omega)$ but from the convergence field $\kappa(\Omega)$, defined as

$$\kappa = -\frac{1}{2}\nabla^2\phi. \quad (7.17)$$

Convergence fields were generated using NaMaster (Alonso, Sanchez & Slosar 2019), from power spectra generated using the Core Cosmology Library⁴ (Chisari et al. 2019). Unlensed CMB fields were also generated using NaMaster, from temperature anisotropy power spectra generated using CAMB⁵ (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012). The implementation of Equation (7.16) initially used the LensTools⁶ code (Petri 2016); however, due to issues arising from dependencies within LensTools, the lensing functionality was later extracted out into a custom Python module. The effects of a telescope beam and detector noise were neglected, as was the non-Gaussian nature of the true convergence field.

For the early stages of this project (described in Sections 7.4.1–7.4.3), lensing was exaggerated to simplify the problem. This was achieved by multiplying each convergence map by a constant factor (greater than 1) prior to simulating lensing. An example of the lensing process using a $30\times$ exaggeration is shown in Figure 7.2. The left panel shows the unlensed CMB field, the centre panel shows the convergence field (limited to $\ell_{\text{max}} = 50$), and the right panel shows the lensed CMB field. The lensing effect is clearly visible, which would not be the case without the $30\times$ exaggeration.

⁴ <https://github.com/LSSTDESC/CCL>

⁵ <https://camb.info>

⁶ <https://lenstools.readthedocs.io/en/latest>

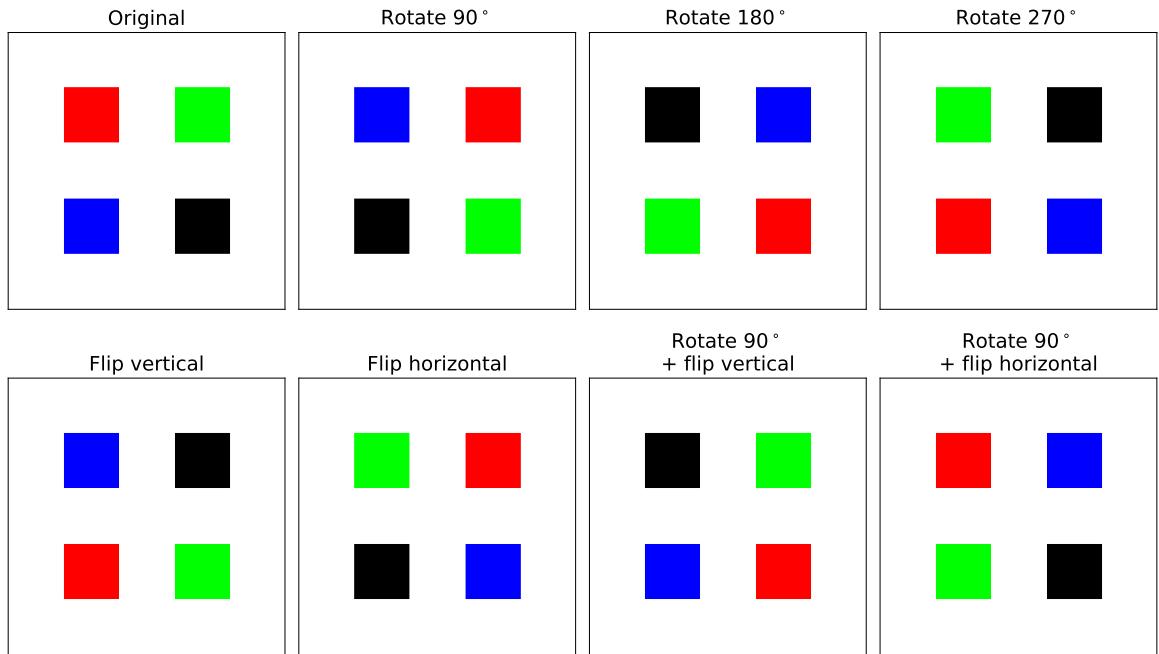


Figure 7.3. Demonstration of the eight unique transforms used for data augmentation, each formed by a different combination of a rotation and a reflection (flip).

From complexity level 2 onwards (described in Sections 7.4.2–7.4.6), a method of data augmentation was used to multiply the size of the training set by a factor 8 without the need to simulate additional realisations. This also had the benefit of implicitly teaching the rotational invariance of lensing to the models during training. The augmentation method used eight unique combinations of rotations and reflections (flips). These were implemented using NumPy (Harris et al. 2020) functions and are demonstrated in Figure 7.3.

Training, validation and test data were always generated together and identically. Samples were shuffled prior to splitting into training, validation and test sets. Only the training set underwent data augmentation, after which it was re-shuffled prior to training, and again between each training epoch. All samples were scaled consistently to ensure that every pixel had a value between -1 and +1 while approximately maximising contrast within this range. This scaling was carried out independently for each complexity level.

7.3.2. Model building and training

Convolutional neural networks were built, trained and tested using the open-source software Keras⁷ (Chollet et al. 2015). Keras is a Python deep learning API for TensorFlow⁸ (Abadi et al. 2016), which is a powerful open-source machine learning library developed by Google. TensorFlow is optimised for the linear algebra operations that are inherent in

⁷ <https://keras.io>

⁸ <https://www.tensorflow.org>

neural network training. It is possible to develop neural networks directly in TensorFlow without using the Keras interface, but Keras greatly speeds up development by offering a simple yet highly customisable API.

The details of the neural network architectures used, and of the training process, are given in Section 7.4. These were developed and explored empirically using an iterative process of trial and error, bounded by the practical constraints of finite time and computing resources. Most of the time, it was possible to hold the entire training set in memory during training, even with the size of the training set being multiplied by 8 by the use of the data augmentation methods described in Section 7.3.1 above. However, for the final level of complexity explored in this chapter, which is described in Section 7.4.6, the larger image size used (100×100 pixels) meant this approach was no longer feasible. An alternative pipeline was therefore developed using a custom subclass of the Sequence object in Keras. The minimum functionality required of such a class is to provide a single batch of training data on request. In the default case, all batches would be loaded into memory on the initialisation of the object, and each would be supplied in turn on request. However, since this would take up too much memory in the case described here, the training set was not preloaded into memory, and instead each batch was loaded from disk only as it was requested, before being removed from memory when the subsequent batch was requested. This allowed the memory requirements to be significantly reduced when the volume of training data was large.

7.4. Models explored and results with increasing complexity

As mentioned above, the structure of this project was to start with a relatively simple (and therefore unrealistic) problem, and gradually add complexity to converge towards a more realistic case. Six increasing levels of complexity are presented in this section, which are summarised below.

1. 22.9 degree field of view covered by 50 pixels along each side (equivalent HEALPix resolution: $N_{\text{side}} = 128$); CMB $\ell_{\text{max}} = 383$; convergence $\ell_{\text{max}} = 50$; convergence exaggerated by factor 30; same unlensed CMB realisation used throughout training, validation and testing.
2. As level 1, except a different CMB realisation used for each training, validation and test sample.
3. As level 2, but convergence exaggeration reduced to a factor 5.
4. As level 3, but no convergence exaggeration.
5. As level 4, except field of view reduced to 21.5 arcmin (equivalent HEALPix resolution: $N_{\text{side}} = 8192$); no CMB ℓ_{max} imposed; convergence $\ell_{\text{max}} = 24\,575$.

6. As level 4, except field of view reduced to 10 degrees and number of pixels along each side increased to 100 (approximate equivalent HEALPix resolution: $N_{\text{side}} \approx 512$); no ℓ_{max} imposed for CMB or convergence. (Note that while all other complexity levels are each based on the previous level, levels 5 and 6 are alternative progressions from level 4.)

In the following subsections, each complexity level is described further, with a description of the different convolutional neural network architectures explored and the results obtained.

7.4.1. Complexity level 1: Same CMB realisation, different lensing realisations

The investigation began with an extremely simple case of a single realisation of the CMB temperature field being lensed by many different convergence realisations. Each convergence field realisation was different for every training, validation, and test sample, but the unlensed CMB map used in each case was identical. The question was whether a CNN could learn the mapping from lensed CMB map to convergence map. To increase its chances, the lensing was strongly exaggerated by multiplying each convergence realisation by a factor 30 at the map level prior to applying the lensing transform, when generating the training data following the method described in Section 7.3.1. A wide field covering a square of side 22.9 degrees was used, with a very low resolution of 50 pixels along each side. This was chosen to be equivalent in pixel area to a HEALPix (Górski et al. 2005) resolution of $N_{\text{side}} = 128$. The input CMB power spectrum was limited by the resolution of the map to a maximum multipole of $\ell_{\text{max}} = 383$, while the convergence power spectrum was truncated at $\ell_{\text{max}} = 50$ to further simplify the scenario.

Relatively few examples of image-to-image CNN models exist in the literature—excluding generative models, which are inappropriate for this problem as here we do not just require a generic realistic convergence map; we require the particular convergence map that corresponds to a specific lensed CMB map.⁹ One example of an image-to-image CNN model that appeared suitable as a starting point was the image super-resolution model described in Shi et al. (2016) and implemented as a Keras code example by Long (2020). The model is designed to be able to scale a previously unseen low-resolution image (or video) to higher resolution, using information it had learned from previously seeing many low- and high-resolution pairs of images in training. While this may seem distant from the problem at hand, it provides a valuable working example of an image-to-image CNN pipeline to use as a baseline. The model was first stripped of its super-resolution functionality, which was provided by a single convolutional layer with a low-resolution input and a higher-resolution output.

⁹ A notable exception is the DeepCMB model of Caldeira et al. (2019), which has similar aims to this work, but we were not aware of that paper at the time.

The remaining model used four convolutional layers: 64 nodes with a 5×5 pixel kernel, 64 nodes with a 3×3 kernel, 32 nodes with a 3×3 kernel, and 1 final node with a 3×3 kernel. All convolutions used a padding method (named `same` in Keras) whereby each image is padded with zeros prior to convolution such that the output from each layer has the same dimensions as the input. Each layer used orthogonal initialisation and a ReLU activation function (see Section 7.2.2). All other options in the model followed the Keras defaults: single-stride convolution was used, with a bias vector initialised as zeros; no regulariser was applied to any kernel, bias vector, or output, nor was a constraint function applied to any kernel or bias vector.

For training, the Adam optimiser was used (see Section 7.2.2) with the default parameters of a learning rate of $\alpha = 10^{-3}$, exponential decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and numerical stability constant $\epsilon = 10^{-7}$. The AMSGrad variation of the Adam algorithm introduced in Reddi, Kale & Kumar (2018), which may assist convergence in some cases, was not used. A mean squared error loss function was used (Equation 7.9).

Many variations from this baseline model were explored. Each was trained for 5 epochs with a training set of 800 samples and a batch size of 32, and validated once per epoch with a separate validation set of 200 samples, and finally evaluated against an unseen test set of 3 samples. Evaluation was primarily a visual comparison to the true convergence field for each test sample (an example may be seen for the best-performing model in Figure 7.4, which is described below). The variations explored were as follows:

- Changing the kernel initialisation from orthogonal to the Glorot uniform initialiser (Glorot & Bengio 2010), which is the Keras default for a convolutional layer. This draws the initial value for each kernel pixel W independently from a uniform distribution as

$$W \sim U(-W_{\max}, W_{\max}), \quad (7.18)$$

with

$$W_{\max} = \sqrt{\frac{6}{n_j + n_{j+1}}}, \quad (7.19)$$

where n_j is the number of nodes in layer j . However, the performance with this method was slightly worse compared to orthogonal initialisation.

- Varying the kernel size in the final layer: sizes of 3×3 , 5×5 , 7×7 , 9×9 and 11×11 pixels were explored, with performance peaking at 9×9 for this particular low-resolution setup.
- Varying the kernel size in the other layers: a small kernel of size 3×3 was found to be optimal.
- Adding additional layers: many combinations of number of nodes and kernel size were explored, from 2 nodes to 64 and from 3×3 to 7×7 , but each one degraded

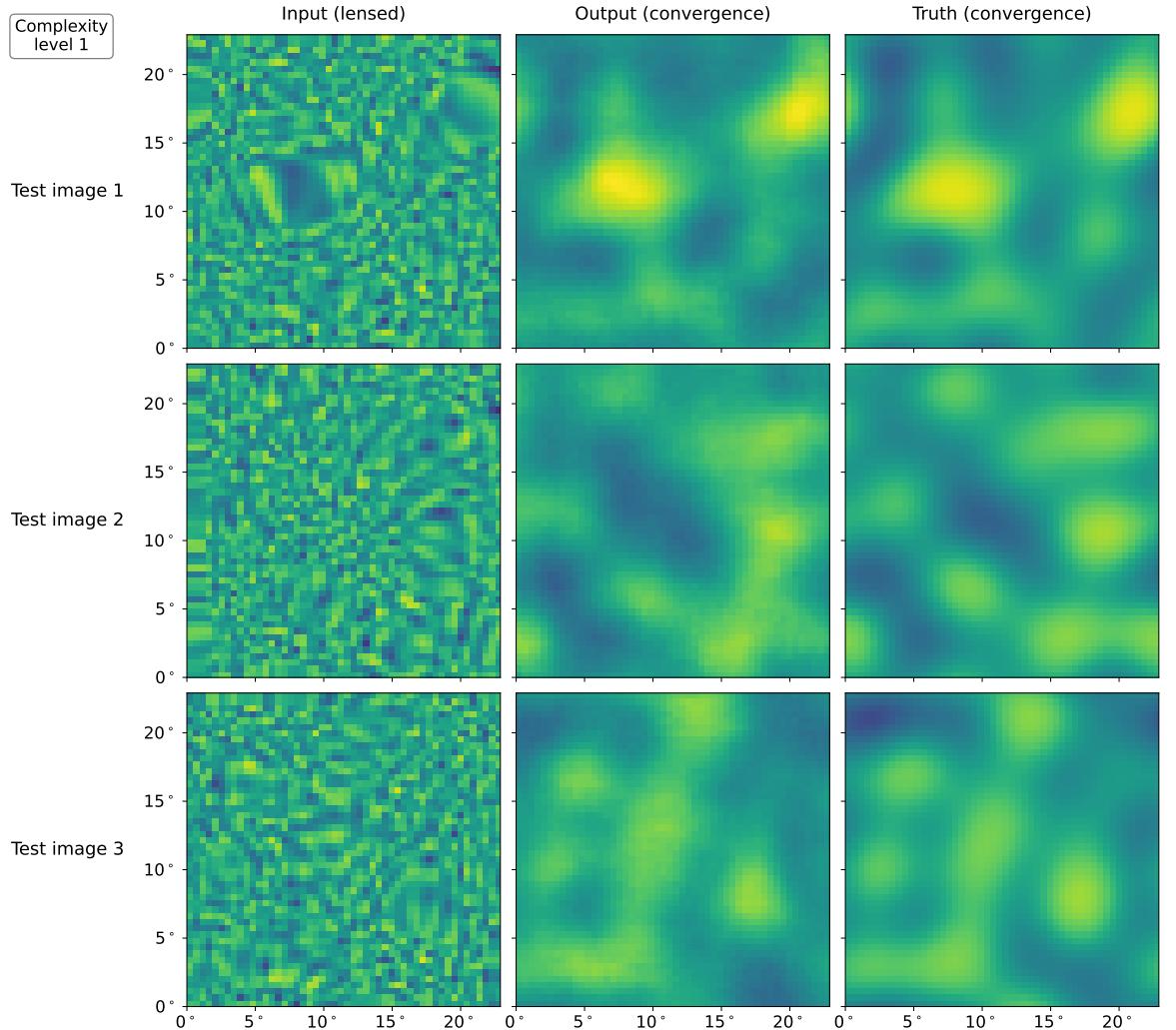


Figure 7.4. Performance of the best-performing model for complexity level 1 (described in Section 7.4.1) on three previously unseen test images. Rows correspond to the different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model, and the true convergence map that was used to generate the test image in the first column. The model was trained for 50 epochs on a training set of 800 samples, with a validation set of 200 samples.

performance compared to the baseline model.

- Removing layers, but removing any layer was found to degrade the performance of the model.
- Increasing the number of nodes in the first layer from 64 to 128, which also degraded the performance, so increasing the numbers of nodes in other layers was not explored.

The conclusion from all of this testing was that the best-performing model over 5 epochs was only a very small variation on the initial model taken from the image super-resolution problem: four convolutional layers comprising 64 nodes with a 3×3 kernel, another 64 nodes with a 3×3 kernel, 32 nodes with a 3×3 kernel, and 1 final node with a 9×9 kernel. This model was then trained for 50 epochs using the same training and validation sets as described above (800 training samples and 200 validation samples), reaching a minimum validation loss on epoch of 45 of $\mathcal{L}_{\text{val}} = 2.1 \times 10^3$.¹⁰ The model was then applied to an unseen test set of three samples, which are shown in Figure 7.4. The three rows correspond to the three different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model, and the true convergence map that was used to generate the test image in the first column. Visually the performance is very good, with a clear correspondence between the estimate and the truth for all three test samples.

7.4.2. Complexity level 2: Different CMB realisations

The second level of complexity was to no longer use the same unlensed CMB realisation across all samples; instead, a different realisation (from the same underlying power spectrum) was used for every sample in the training, validation and test sets. All other aspects were identical to complexity level 1 described in Section 7.4.1 above, including that each convergence field was exaggerated by a factor 30.

The baseline model for this level was a small variation on the best model from the previous level. Early testing prior to implementing data augmentation (described below) yielded a better performance with 64 nodes on the third layer as well as the first two, giving a baseline model comprising four convolutional layers: three of 64 nodes with a 3×3 kernel followed by 1 final node with a 9×9 kernel. All other aspects of the model, optimiser and loss function used were identical to complexity level 1. This model will be referred to as L2_baseline.

For this complexity level, data augmentation was introduced to multiply the size of the training set by a factor 8, as described in Section 7.3.1, as well as to implicitly teach the rota-

¹⁰ Quoted loss values cannot be compared between different complexity levels, because each level uses its own scaling of the training, validation and test samples to ensure that all inputs to the CNN are in the range (-1, 1) while approximately maximising the contrast within this range, as discussed in Section 7.3.1.

tional invariance of lensing to the model. Prior to augmentation, the training set contained 800 samples and the validation set contained 200 samples. Augmentation was applied to the training set to give 6400 training samples. The validation set was not augmented to avoid any possible bias.

Data augmentation turned out to be highly effective in increasing the performance of the model over a given number of epochs. Figure 7.5 shows the L2_baseline model evaluated on an unseen test set (which was not augmented) after having been trained for 10 epochs on the same training set with and without data augmentation. There is clearly a closer resemblance to the true convergence field in the augmented case. This is also illustrated by the training and validation loss per training epoch in each case, which is shown in Figure 7.6. There is a consistent factor 2.5–5 improvement in both loss values throughout training, reaching respective minima of $\mathcal{L}_{\text{val}} = 5.9 \times 10^3$ without augmentation and $\mathcal{L}_{\text{val}} = 2.2 \times 10^3$ with augmentation.

With data augmentation in place, three variations on the baseline model were explored, which are summarised along with the baseline model as follows:

- L2_baseline:
Layers 1–3: 64 nodes with a 3×3 kernel;
Layer 4: 1 node with a 9×9 kernel.
- L2_+32:
As L2_baseline, but with an additional layer containing 32 nodes with a 3×3 kernel before the final layer.
- L2_+64:
As L2_baseline, but with an additional layer containing 64 nodes with a 3×3 kernel before the final layer.
- L2_first-layer-128:
As L2_baseline, but with the first layer having 128 nodes instead of 64.

Each model was trained for 5 epochs on the same training set, containing 800 samples prior to augmentation, which increased the number of samples by a factor 8. A validation set containing 200 samples was again used, and not augmented. Each model was evaluated based on its visual performance on a test set of 3 previously unseen images, and on its validation loss during training. These two measures were consistently in agreement. Figure 7.7 shows the validation loss of each model over the training period. The best-performing models were the L2_baseline and L2_+64 models, which both ended epoch 5 with an identical validation loss of $\mathcal{L}_{\text{val}} = 2.6 \times 10^{-3}$.

The L2_+64 model was re-trained for 10 epochs to obtain a comparison with the L2_baseline model, which was previously trained for 10 epochs when testing the effect of data augmentation as described above. This revealed a visual performance on the test set

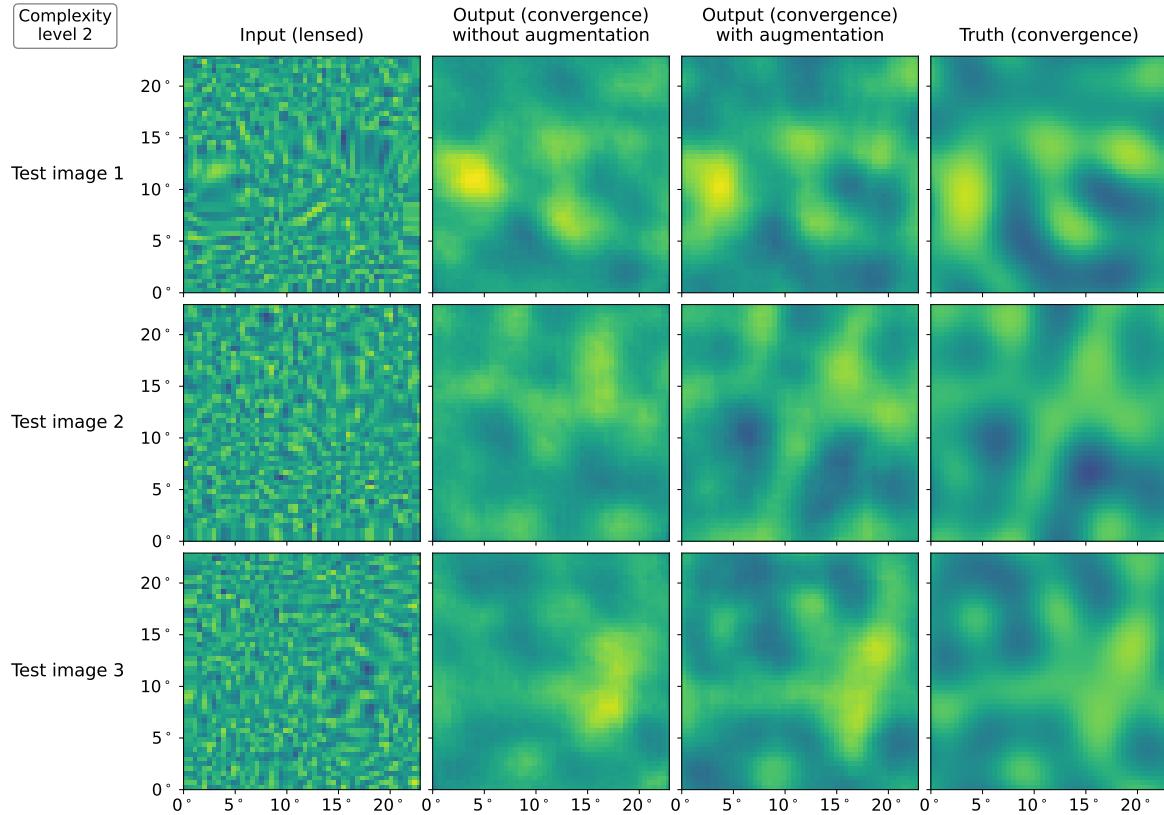


Figure 7.5. The effect of data augmentation on the performance of the L2_baseline model, evaluated on three previously unseen test samples after having been trained for 10 epochs on the same training set with and without data augmentation. Rows correspond to the different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model after training without data augmentation, the predicted convergence map after training with data augmentation, and the true convergence map that was used to generate the test image in the first column. The training set contained 800 samples prior to augmentation, with a validation set of 200 samples which was not augmented.

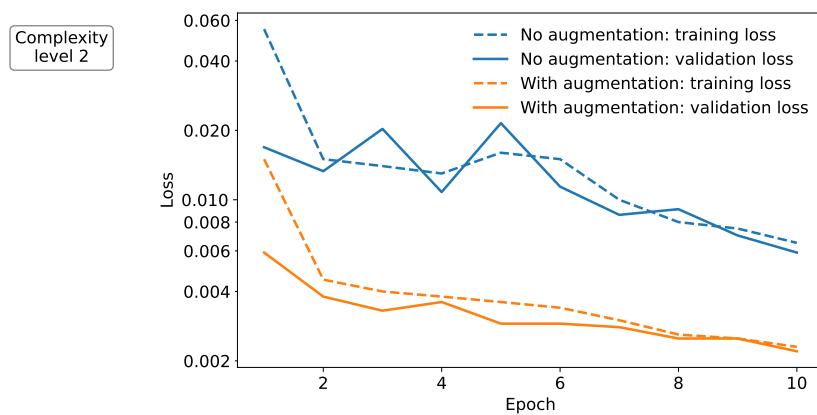


Figure 7.6. Training and validation loss for the L2_baseline model over 10 epochs with and without data augmentation. The training set contained 800 samples prior to augmentation, with a validation set of 200 samples which was not augmented.

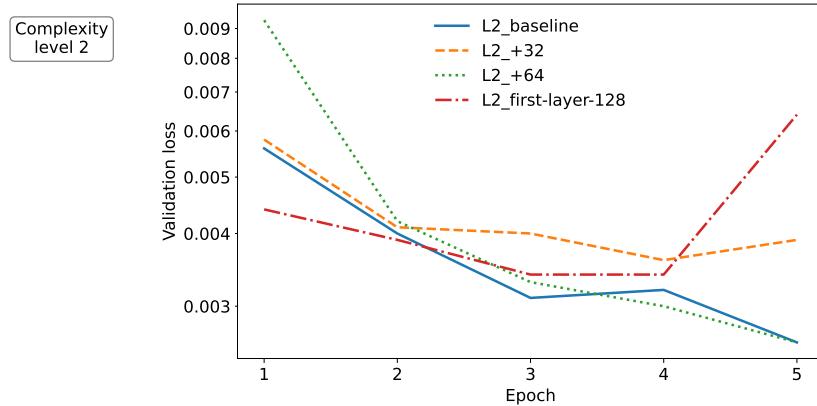


Figure 7.7. Validation loss compared between the four models explored in complexity level 2, which are described in Section 7.4.2. Each model was trained for 5 epochs on the same training set containing 800 samples prior to augmentation by a factor 8, and validated on 200 different samples which were not augmented.

that slightly surpassed that of the baseline model, with a validation loss of $\mathcal{L}_{\text{val}} = 1.8 \times 10^{-3}$ (compared to the corresponding 10-epoch value obtained above for the L2_baseline model of $\mathcal{L}_{\text{val}} = 2.2 \times 10^{-3}$). The L2_+64 model was therefore selected to train for 50 epochs on the same training set, reaching a best validation loss of $\mathcal{L}_{\text{val}} = 1.13 \times 10^{-3}$ in epoch 50. The predictions made by this model on the previously unseen test set are shown in Figure 7.8. Visually the performance is good, roughly comparable with those from complexity level 1 (Figure 7.4) despite using a different unlensed CMB realisation for each sample.

7.4.3. Complexity level 3: Reduced lensing exaggeration

The third level of complexity was to reduce the exaggeration of the convergence field to a factor 5, from the value of 30 used in levels 1 and 2. All other aspects of the setup were identical to level 2, i.e. a different unlensed CMB realisation was used for each sample and the resolution settings were the same as described for level 1 in Section 7.4.1.

The exaggeration factor of 5 was chosen as the result of some initial exploratory testing, carried out over 5 epochs with a training set containing 800 samples, augmented by a factor 8, and a validation set of 200 samples, which was not augmented. These tests used the best-performing L2_+64 model from complexity level 2 (described in Section 7.4.2 above) as the new baseline model, here renamed L3_baseline. This initial testing found that disabling lensing exaggeration entirely yielded a completely blank prediction of the convergence field for all test samples, and a validation loss that immediately plateaued at $\mathcal{L}_{\text{val}} = 1.4 \times 10^{-2}$. This would be a difficult starting point, since a working baseline is effectively an essential requirement for finding iterative improvements. An exaggeration factor of 10 was also explored, but this performed comparably well to a factor of 30, reaching a 5-epoch validation loss of $\mathcal{L}_{\text{val}} = 2.8 \times 10^{-3}$, compared to $\mathcal{L}_{\text{val}} = 2.0 \times 10^{-3}$ for an equivalent run

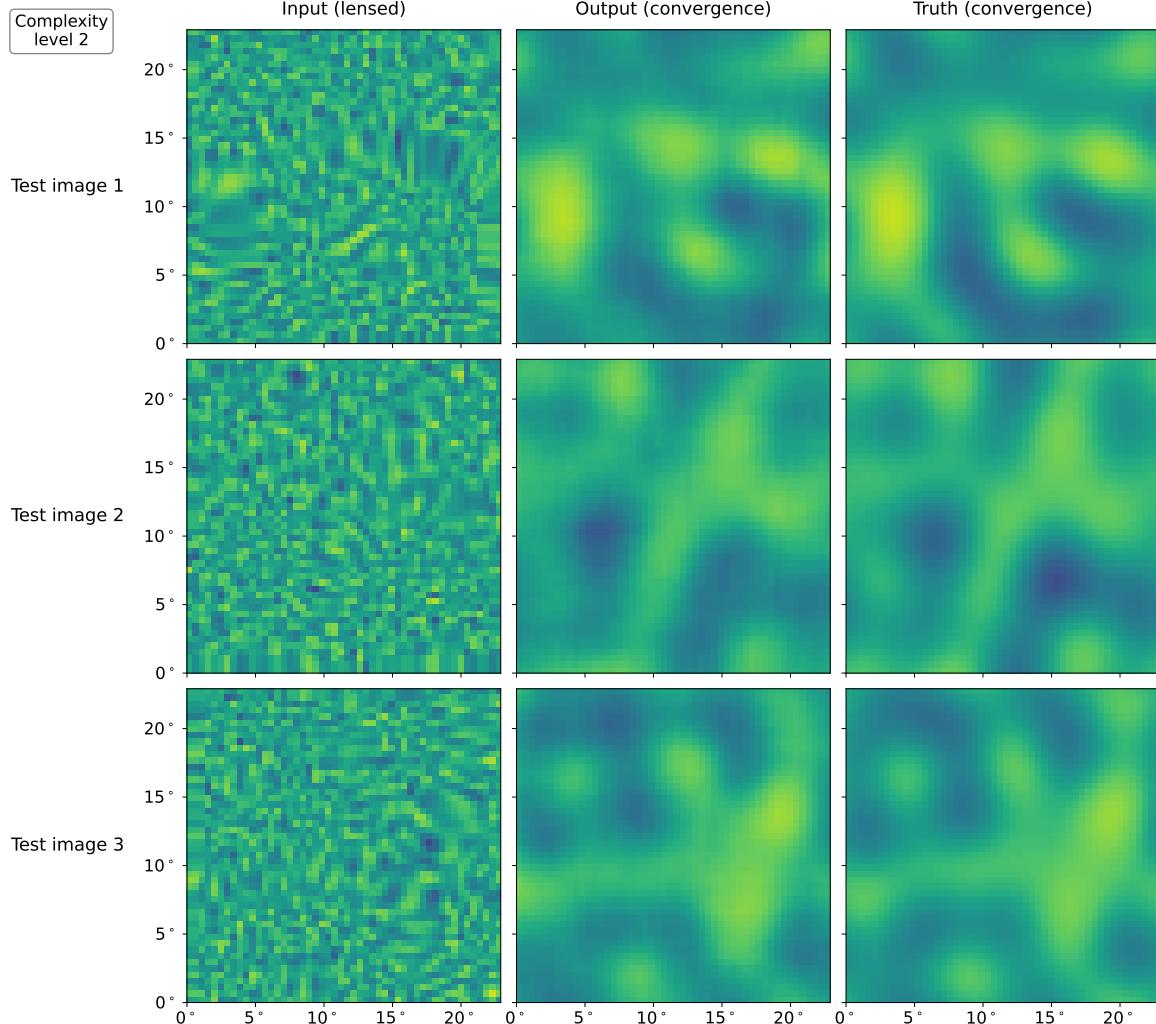


Figure 7.8. Performance of the best-performing L2_+64 model for complexity level 2 (described in Section 7.4.1) on three previously unseen test images. Rows correspond to the different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model, and the true convergence map that was used to generate the test image in the first column. The model was trained for 50 epochs on a training set of 800 samples augmented by a factor 8, with a validation set of 200 samples which was not augmented.

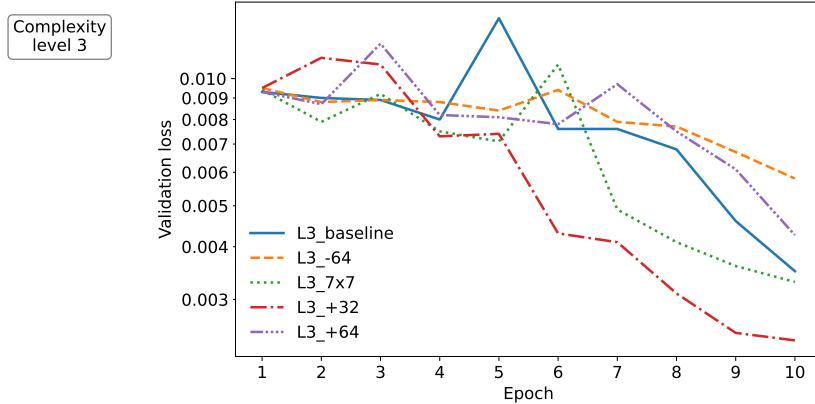


Figure 7.9. Validation loss compared between the five best-performing models explored in complexity level 3, which are described in Section 7.4.3. Each model was trained for 10 epochs on the same training set containing 800 samples prior to augmentation by a factor 8, and validated on 200 different samples which were not augmented.

with data from complexity level 2 (using an equivalent scaling, therefore allowing for a valid comparison). An exaggeration factor of 5 was ultimately chosen as it produced a relatively poor performance ($\mathcal{L}_{\text{val}} = 8.2 \times 10^{-3}$) in the L3_baseline model over 5 epochs, while still producing a validation loss that was steadily decreasing and being able to predict some features in the true convergence maps of the test sample. This made it a suitable starting point for improvements, with an aim towards removing the lensing exaggeration altogether in the next complexity level (Section 7.4.4).

As mentioned above, the baseline model for this complexity level was identical to the best-performing L2_+64 model from complexity level 2, including all aspects of the training setup. Several variations on this model were explored. The standout performers in initial 5-epoch tests were selected for longer 10-epoch tests, which are described below. These standout performers, alongside the baseline model, are as follows:

- L3_baseline:
Layers 1–4: 64 nodes with a 3×3 kernel;
Layer 5: 1 node with a 9×9 kernel.
- L3_-64:
As L3_baseline, but without one of the 64-node layers. (This is identical to the L2_baseline model.)
- L3_7x7:
As L3_baseline, but with the final lone kernel having size 7×7 .
- L3_+32:
As L3_baseline, but with an additional layer of 32 nodes with a 3×3 kernel before the final layer.

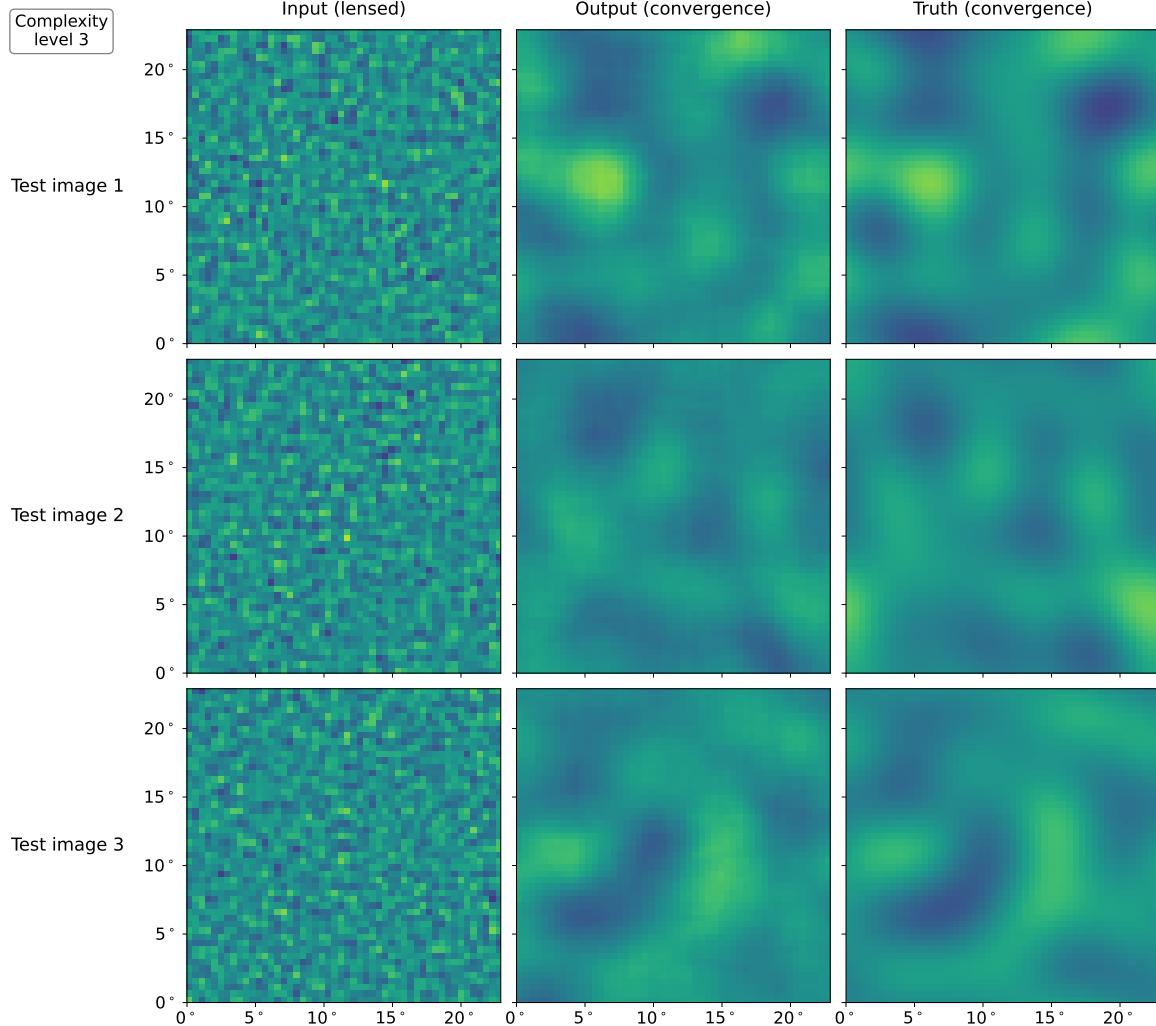


Figure 7.10. Performance of the best-performing L3_+32 model for complexity level 3 (described in Section 7.4.3) on three previously unseen test images. Rows correspond to the different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model, and the true convergence map that was used to generate the test image in the first column. The model was trained for 50 epochs on a training set of 800 samples augmented by a factor 8, with a validation set of 200 samples which was not augmented.

- L3_+64:

As L3_baseline, but with an additional layer of 64 nodes with a 3×3 kernel before the final layer.

The other variations that were explored in the 5-epoch tests included adding and removing different layers, changing numbers of nodes, and changing kernel sizes, but other than the models listed above all variations were found to be either inferior to or similar to the baseline model.

The five models listed above were each trained for 10 epochs using the same training and validation sets as described above (training size: 800×8 ; validation size: 200). The validation loss of each model for each training epoch is shown in Figure 7.9. There is a clear victory for the L3_+32 model, which achieved a best validation loss of $\mathcal{L}_{\text{val}} = 2.4 \times 10^{-3}$, while the L3_-64 ($\mathcal{L}_{\text{val}} = 5.8 \times 10^{-3}$) and L3_+64 ($\mathcal{L}_{\text{val}} = 4.3 \times 10^{-3}$) models performed worse than the baseline ($\mathcal{L}_{\text{val}} = 3.5 \times 10^{-3}$). The L3_7x7 model achieved a slight improvement on the baseline, of $\mathcal{L}_{\text{val}} = 3.3 \times 10^{-3}$. These respective performances were also mirrored in visual tests on a previously unseen test set containing 3 new samples.

The L3_+32 model was subsequently trained for 50 epochs using the same training and validation set, reaching a best validation loss of $\mathcal{L}_{\text{val}} = 1.173 \times 10^{-3}$. It was then evaluated on the unseen test set of 3 samples. The resulting predictions for the convergence field are shown alongside the lensed CMB maps and true convergence fields in Figure 7.10. The model achieves a good performance, comparable to complexity level 2 (Figure 7.8), despite the reduction in the lensing exaggeration, which is clearly visible in the left column of Figure 7.10 when compared to the visibly exaggerated lensing effect in previous complexity levels (Figures 7.4 and 7.8). This comparable level of performance to complexity level 2 has been achieved without a larger amount of data or more training, and simply required a slightly more complex model architecture.

7.4.4. Complexity level 4: No lensing exaggeration

With a working solution in place for the case of lensing exaggerated by a factor 5, the fourth level of complexity was to now remove the exaggeration entirely. All other aspects of the setup were identical to complexity levels 2–4, which were described in Section 7.4.2.

The best-performing L3_+32 model from complexity level 3 (Section 7.4.3) was adopted as the baseline model for level 4, here renamed L4_baseline. Early testing over 10 epochs with a training set of 800 samples, augmented by a factor 8 following the methods described in Section 7.3.1, failed to recover any features in the true convergence maps. This training set was subsequently enlarged to 1600 samples, but no improvement was detected. As a result, the decision was made to train three models with different depths over a large number of epochs with a large training set. The aim of this approach was twofold: if a

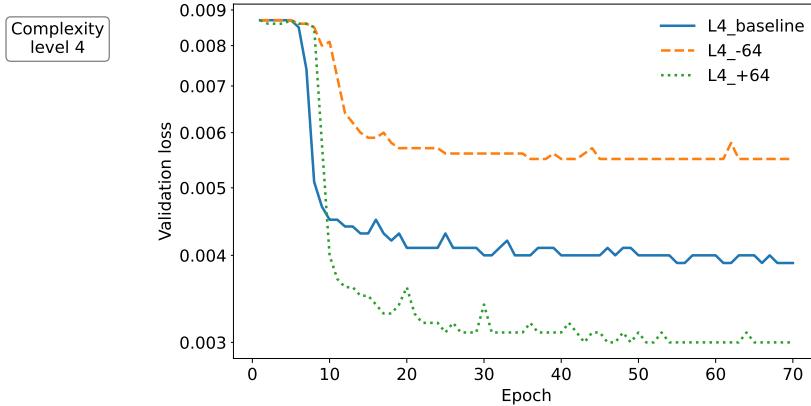


Figure 7.11. Validation loss compared between the three models explored in complexity level 4, which are described in Section 7.4.4. Each model was trained for 70 epochs on the same training set containing 9800 samples prior to augmentation by a factor 8, and validated on 200 different samples which were not augmented.

large amount of training did not help, it would imply that the lack of data was not the problem and therefore that the models were in some way inappropriate, while if there was a discernible difference between the performance of the three models, this could point to a natural direction in which to seek further improvement.

The three chosen models are as follows:

- L4_baseline:
 - Layers 1–4: 64 nodes with a 3×3 kernel;
 - Layer 5: 32 nodes with a 3×3 kernel;
 - Layer 6: 1 node with a 9×9 kernel.
- L4_-64:
 - As L4_baseline, but without one of the 64-node layers.
- L4_+64:
 - As L4_baseline, but with an additional 64-node layer with a 3×3 kernel before the final layer.

Each model was trained for 70 epochs on a training set containing 9800 unique samples, augmented by a factor 8 to a final size of 78 400. A separate validation set of 200 samples was used, which was not augmented. The validation loss during training for each model is shown in Figure 7.11, which clearly indicates that more model complexity is beneficial: the more complex L4_+64 model achieved a validation loss of $\mathcal{L}_{\text{val}} = 2.98 \times 10^{-3}$, while the L4_baseline model reached $\mathcal{L}_{\text{val}} = 3.91 \times 10^{-3}$ and the simpler L4_-64 only reached $\mathcal{L}_{\text{val}} = 5.45 \times 10^{-3}$. All three models converged to a stable validation loss after around 15–30 epochs of training, indicating that they had exhausted the information available in the finite training set.

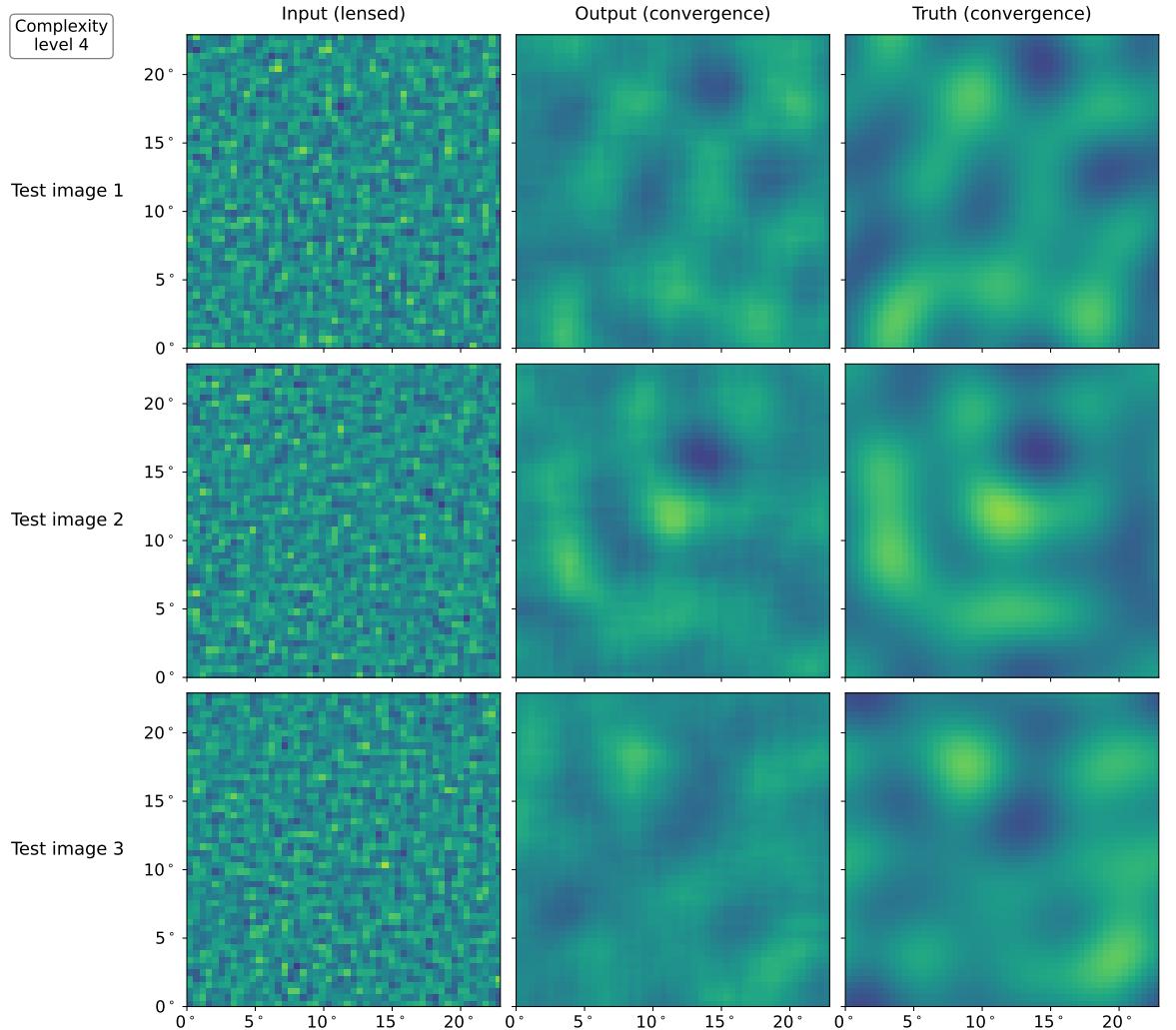


Figure 7.12. Performance of the best-performing L4_+64 model for complexity level 4 (described in Section 7.4.4) on three previously unseen test images. Rows correspond to the different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model, and the true convergence map that was used to generate the test image in the first column. The model was trained for 70 epochs on a training set of 9800 samples augmented by a factor 8, with a validation set of 200 samples which was not augmented.

After training, the three models were each evaluated on a previously unseen test set containing three samples. The results were consistent with the respective validation loss values for each model (Figure 7.11), and the L4_+64 was clearly the best performer in terms of the visual correspondence of its predictions to the true convergence field. This is shown in Figure 7.12. The performance is reasonable, although clearly worse than previous complexity levels despite the extra volume of training—both in the size of the training set and the number of training epochs. There are some signs of potential artefacts in the predicted convergence maps. Similar features were observed previously in early tests for the previous complexity levels, but in those cases such features always disappeared with additional training epochs. In this case, the convergence of the validation loss to a steady model for all three models (seen in Figure 7.11) indicates that these features should not be expected to disappear with still more training epochs. However, there are clear indications of possible ways in which the performance seen in Figure 7.12 might be improved. First, a significant improvement was observed when enlarging the training set from 800 or 1800 samples—which failed to recover any features in the true convergence maps—to 9800 samples, which delivered a much better performance, as seen in Figure 7.12, so it is reasonable to expect that still more data may be able to improve the performance of the three models considered here. Additionally, the comparison between the three models indicates that for this problem a deeper model is more successful, so a natural next step might be to try adding additional layers. However, the level of realism considered at this complexity level is still far below a case that might have practical use, so a progression to the next complexity level takes priority over seeking further improvements to this one.

7.4.5. Complexity level 5: Higher resolution, small field of view

The fifth level of complexity was to add resolution, while maintaining the number of pixels at 50×50 in order to retain computational feasibility within reasonable training timescales and memory requirements, both of which rapidly increase with larger images. The field of view was set to a side length of 21.5 arcmin, which allowed a sub-arcminute pixel size of 0.43 arcmin, equivalent in pixel area to a HEALPix resolution of $n_{\text{side}} = 8192$. The low convergence ℓ_{max} value was raised to a resolution-limited value of $\ell_{\text{max}} = 24\,575$, while the CMB ℓ_{max} constraint was removed entirely, being limited by the signal which is negligible above around $\ell \sim 7000$. Such high resolution was partially motivated by the ability to estimate power spectra, for comparison between power spectra estimated from the true and predicted convergence fields.

The baseline model considered for this complexity level was a small variation on the best-performing L4_+64 model from Section 7.4.4. The single 32-node layer was replaced with a 64-node layer like the other multi-node layers, and the final kernel size was reduced from 9×9 to 3×3 , with an aim of being able to recover smaller-scale features in the convergence field. Four variants on this baseline model were considered, which are summarised along

with the baseline model as follows:

- L5_baseline:
Layers 1–6: 64 nodes with a 3×3 kernel;
Layer 7: 1 node with a 3×3 kernel.
- L5_+64:
As L5_baseline, but with an additional 64-node layer with a 3×3 kernel before the final layer.
- L5_7x7:
As L5_baseline, but with the 3×3 kernel in the final single-node layer replaced with a 7×7 kernel.
- L5_deep:
Layers 1–10: 32 nodes with a 3×3 kernel;
Layer 11: 1 node with a 3×3 kernel.
- L5_upsampled:
As L5_baseline, but upsampling by a factor 2 before the first layer, before applying a mean pooling to downsample to the original resolution before the final layer.

There were various motivations for exploring these four variations on the baseline model. L5_+64 was motivated by the results of complexity level 4 (Section 7.4.4), which were clearly suggestive of additional layers achieving higher performance. L5_7x7 was something of an insurance policy in case the reduction of the final kernel size turned out to be detrimental to performance. L5_deep was motivated by the fact that deeper neural networks are, in general, empirically and theoretically superior to wider networks, since they are better able to model non-linearity in the function that maps from the input to the output (e.g. Safran & Shamir 2016; Lee et al. 2020). Finally, L5_upsampled was inspired by the original image super-resolution model of Shi et al. (2016) used as the starting point for complexity level 1 in Section 7.4.1, which used sub-pixel convolution. It seemed plausible that this technique may allow the model to extract smaller-scale information from the lensed CMB maps in this problem.

Each of the five models was trained for 100 epochs, with a training set containing 9800 unique samples, which was augmented by a factor 8, and a validation set containing 200 samples, which was not augmented. The validation loss for each model throughout training is shown in Figure 7.13. The best-performing model was L5_+64, reaching a minimum validation loss of $\mathcal{L}_{\text{val}} = 6.80 \times 10^{-3}$. Three other models performed similarly well: L5_baseline and L5_deep also reached $\mathcal{L}_{\text{val}} = 6.80 \times 10^{-3}$, while L5_deep was slightly behind at $\mathcal{L}_{\text{val}} = 6.83 \times 10^{-3}$. However, all three of these models exhibited a strange behaviour of a sudden divergence in the validation loss followed by a slow convergence back towards the initial minimum. In the case of the L5_deep model, a second divergence was

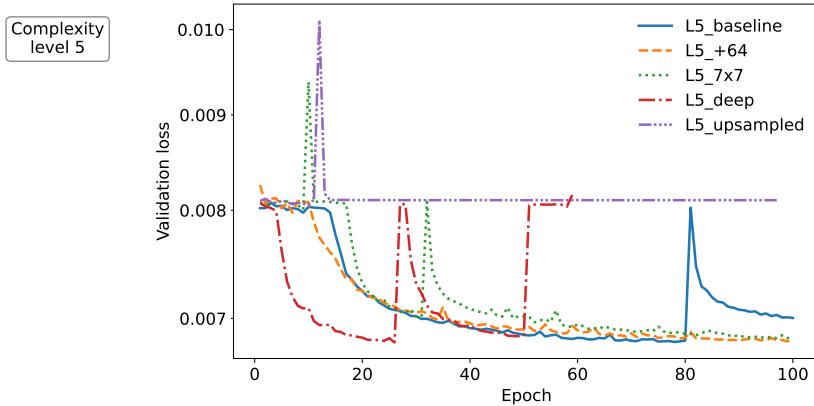


Figure 7.13. Validation loss compared between the five models explored in complexity level 5, which are described in Section 7.4.5. Each model was trained for 100 epochs on the same training set containing 9800 samples prior to augmentation by a factor 8, and validated on 200 different samples which were not augmented.

observed, at which point the model weights became infinite and training ceased; this is why the corresponding line in Figure 7.13 stops suddenly at around 60 epochs. Similar features were found in the training loss. This behaviour is not currently understood, since divergent weights should not be permitted by the update rule for the Adam optimiser (Equation 7.14). This issue is discussed further in Section 7.5. Meanwhile, the performance of the L5_upsampled model was poor, plateauing at $\mathcal{L}_{\text{val}} = 8.05 \times 10^{-3}$.

Each model was evaluated using a test set of three previously unseen images. For each model, the weights corresponding to the minimum validation loss were used—for instance, the L5_baseline model was evaluated using its weights at around epoch 80, prior to its subsequent divergence (seen in Figure 7.13). The results were consistent with the respective validation losses: a similar performance was observed between all models except L5_upsampled, which did not appear to recover any features in the true convergence maps. The visual test results for the L5_+64 model are shown in Figure 7.14. Small-scale features are absent, but there is an apparent correspondence between the estimated and true convergence fields on larger scales.

For the first time, it was possible with this complexity level to estimate power spectra from the true and predicted convergence fields, using NaMaster. These are shown for the three test images and the L5_+64 model in Figure 7.15. It is clear that the model is able to recover large-scale features in the power spectrum quite well, but that small-scale features (above around $\ell \sim 3500$) are missed entirely. This is consistent with the visual observations from the maps in Figure 7.14. Similar performance was found for all other models except L5_upsampled, which did not recover any power.

It is perhaps unsurprising that smaller-scale features in the convergence maps are not recovered, given the lack of visible features on such scales in the input lensed CMB maps (left column of Figure 7.14). This motivated a return to a larger field of view, in order to

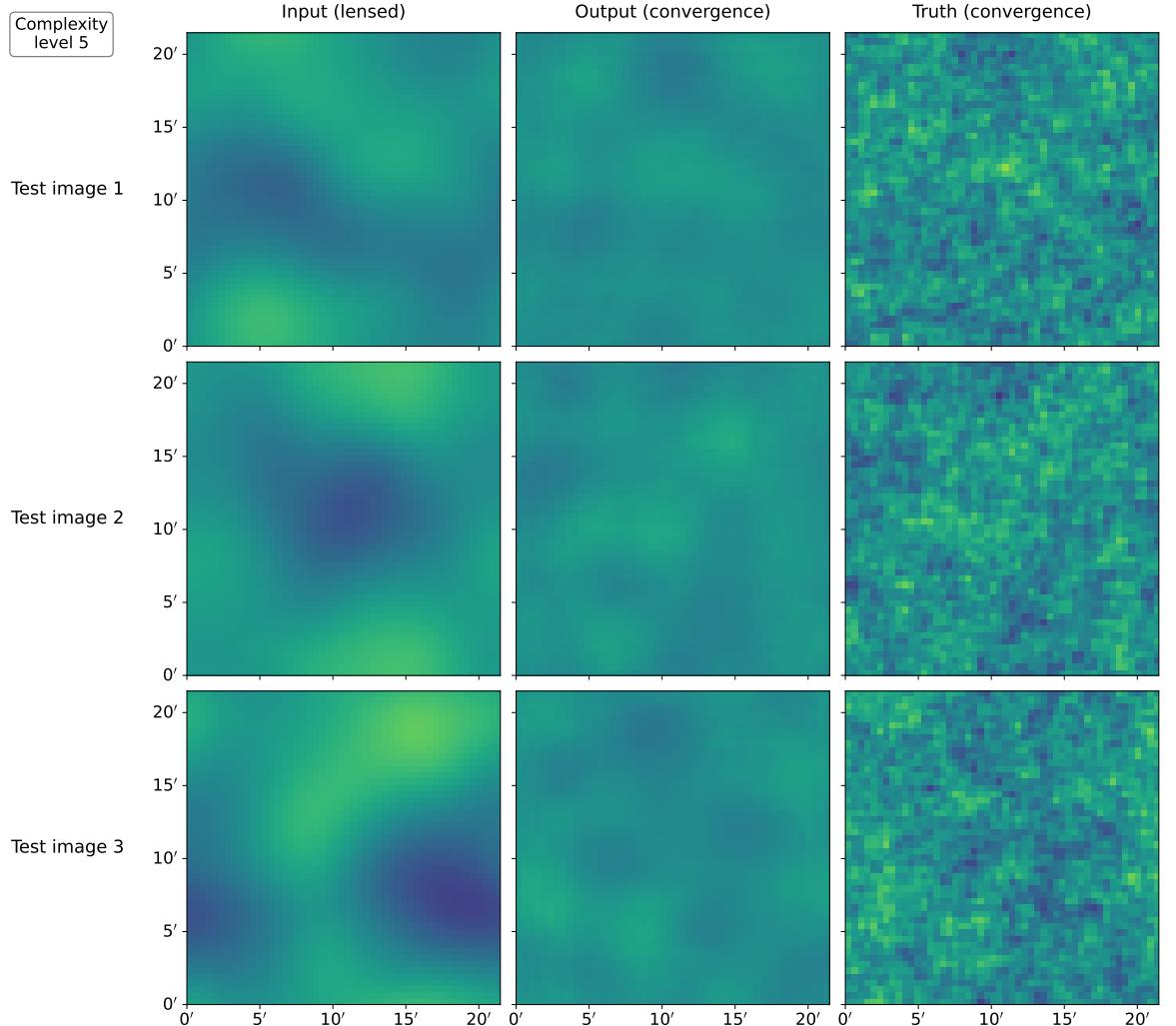


Figure 7.14. Performance of the best-performing L5_+64 model for complexity level 5 (described in Section 7.4.5) on three previously unseen test images. Rows correspond to the different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model, and the true convergence map that was used to generate the test image in the first column. The model was trained for 100 epochs on a training set of 9800 samples augmented by a factor 8, with a validation set of 200 samples which was not augmented.

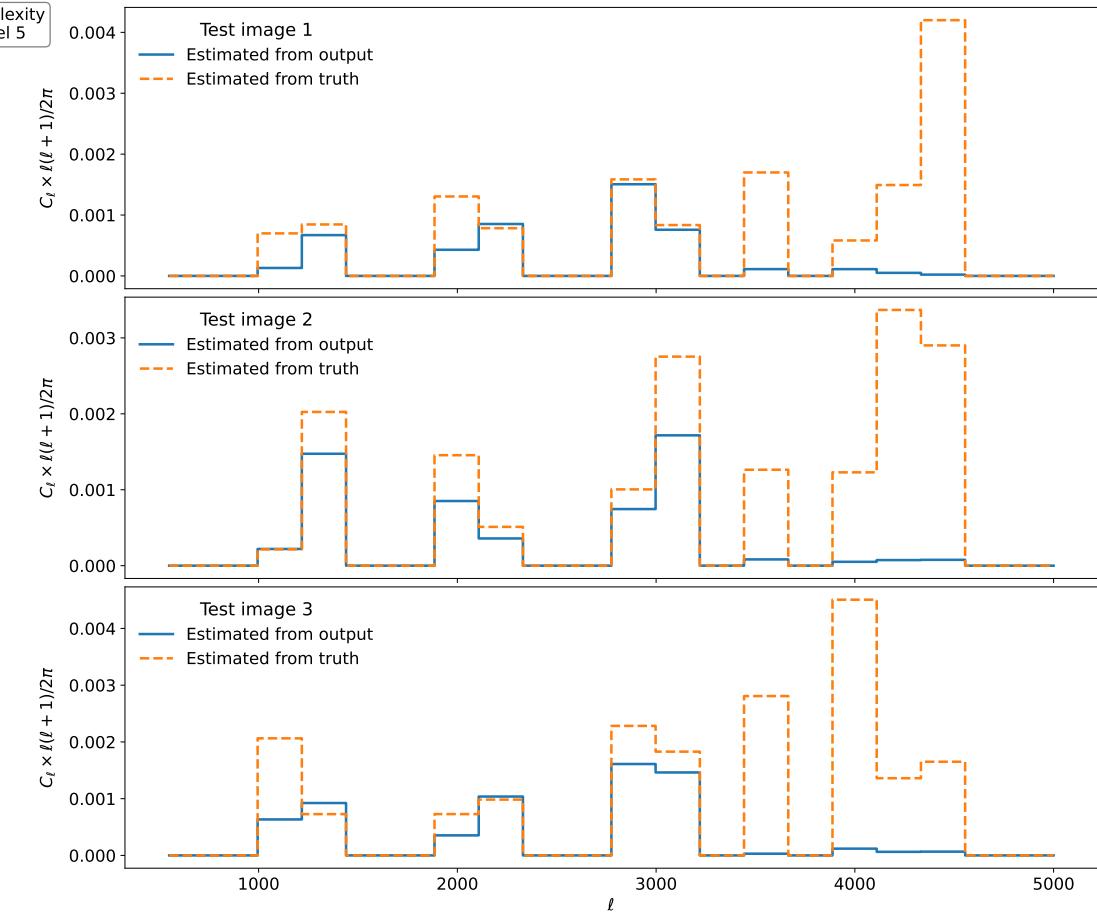


Figure 7.15. Power spectra estimated from the convergence maps shown in Figure 7.14. Rows correspond to the three different test images. The blue solid line shows the power spectrum estimated from the predicted convergence map output by the convolutional neural network, while the orange dashed line shows the power spectrum estimated from the true convergence map. Results shown are for the L5_+64 model for complexity level 5 (described in Section 7.4.5). The model was trained for 100 epochs on a training set of 9800 samples augmented by a factor 8, with a validation set of 200 samples which was not augmented.

regain the degree-scale features in the CMB and convergence fields. This is explored in Section 7.4.6 below.

7.4.6. Complexity level 6: Higher resolution, wide field of view

As mentioned above, for the next complexity level it was chosen to return to a larger field of view, in order to regain the degree-scale features in the CMB and convergence fields. Ideally this would be done while retaining sub-arcminute pixel size, since typical deflections are around this scale (e.g. Lewis & Challinor 2006). However, this would require a large number of pixels, of order 1000 along each side. Initial testing revealed that this was not computationally feasible with the available resources, due to memory requirements during training. As a result, an intermediate setup was instead explored, using 100 pixels along each side with a 10 degree field of view, giving a pixel size of 6 arcmin along each side. Unlike previous complexity levels, there is no exact equivalent HEALPix resolution in terms of pixel area, but it is approximately equivalent to $n_{\text{side}} = 512$. No explicit ℓ_{max} cut was applied to either the CMB or convergence fields, but both were limited by resolution to an effective $\ell_{\text{max}} \sim 1535$.

Even this modest increase of total number of pixels by a factor of 4 (2 along each side) required the introduction of Keras Sequence objects to load only a part of the training data into memory at any given point during training, as described in Section 7.3.2.

Motivated by the ability of the L5_deep model to quickly reach a low minimum validation loss in Section 7.4.5 (the subsequent unexplained divergence notwithstanding), and by the general preference towards deeper models in the literature described in that section, the baseline model was chosen to contain 12 convolutional layers of 32 nodes each, plus the usual final single-node convolutional layer. The final kernel size of 3×3 used in Section 7.4.5 was retained, since the goal of recovering small-scale features is the same. All other aspects of the model, and the optimiser used for training, were identical to previous complexity levels. Two additional model variations with increasing depth and decreasing width were explored. These, along with the baseline model, are summarised as follows:

- L6_baseline:
 - Layers 1–12: 32 nodes with a 3×3 kernel;
 - Layer 13: 1 node with a 3×3 kernel.
- L6_x2depth:
 - Layers 1–24: 16 nodes with a 3×3 kernel;
 - Layer 25: 1 node with a 3×3 kernel.
- L6_x4depth:
 - Layers 1–48: 8 nodes with a 3×3 kernel;
 - Layer 49: 1 node with a 3×3 kernel.

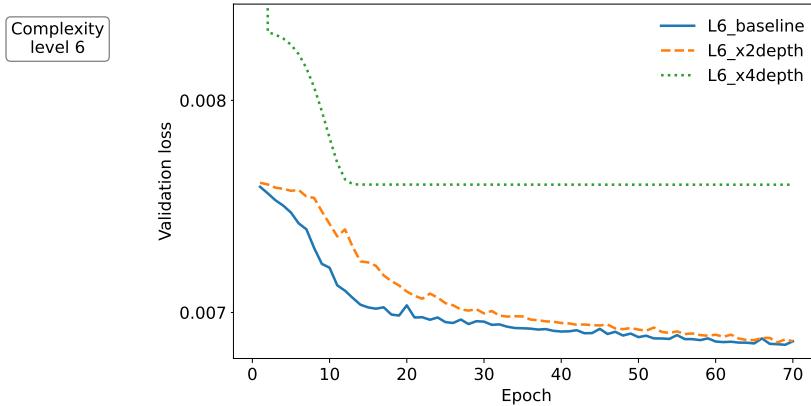


Figure 7.16. Validation loss compared between the three models explored in complexity level 6, which are described in Section 7.4.6. Each model was trained for 70 epochs on the same training set containing 9800 samples prior to augmentation by a factor 8, and validated on 200 different samples which were not augmented.

Each model was trained for 70 epochs on a training set containing 9800 samples, which was augmented by a factor 8 using the methods described in Section 7.3.1, and validated using a validation set containing 200 samples, which was not augmented. The validation loss of each model throughout training is shown in Figure 7.16. The L6_baseline in fact performed the best throughout the whole training process, reaching a validation loss of $\mathcal{L}_{\text{val}} = 6.86 \times 10^{-3}$, although by epoch 70 the performance of the L6_x2depth model was similar, reaching $\mathcal{L}_{\text{val}} = 6.87 \times 10^{-3}$. The L6_x4depth model performed most poorly, with the validation loss plateauing at $\mathcal{L}_{\text{val}} = 7.59 \times 10^{-3}$.

After training, each model was assessed using a test set containing 3 previously unseen images. The resulting estimated convergence map is shown for each image along with the corresponding true convergence map in Figure 7.17, for the best-performing L6_baseline model. It appears that most features are correctly recovered, but with a too-small amplitude. Similar performance was observed for the L6_x2depth model, while the L6_x4depth model did not appear to be able to recover any features in the true convergence maps.

Power spectra were also estimated, and are shown for the L6_baseline model in Figure 7.18. It appears that on larger scales (below $\ell \sim 500$), features are recovered at a reduced amplitude, whereas on smaller scales they are absent entirely, although the reduced amplitude as a function of scale makes a comparison difficult.

This marks the final extent of the project at the time of writing. Possible future directions in which to continue this work are discussed in Section 7.5 below, along with the achievements and challenges of the work to date.

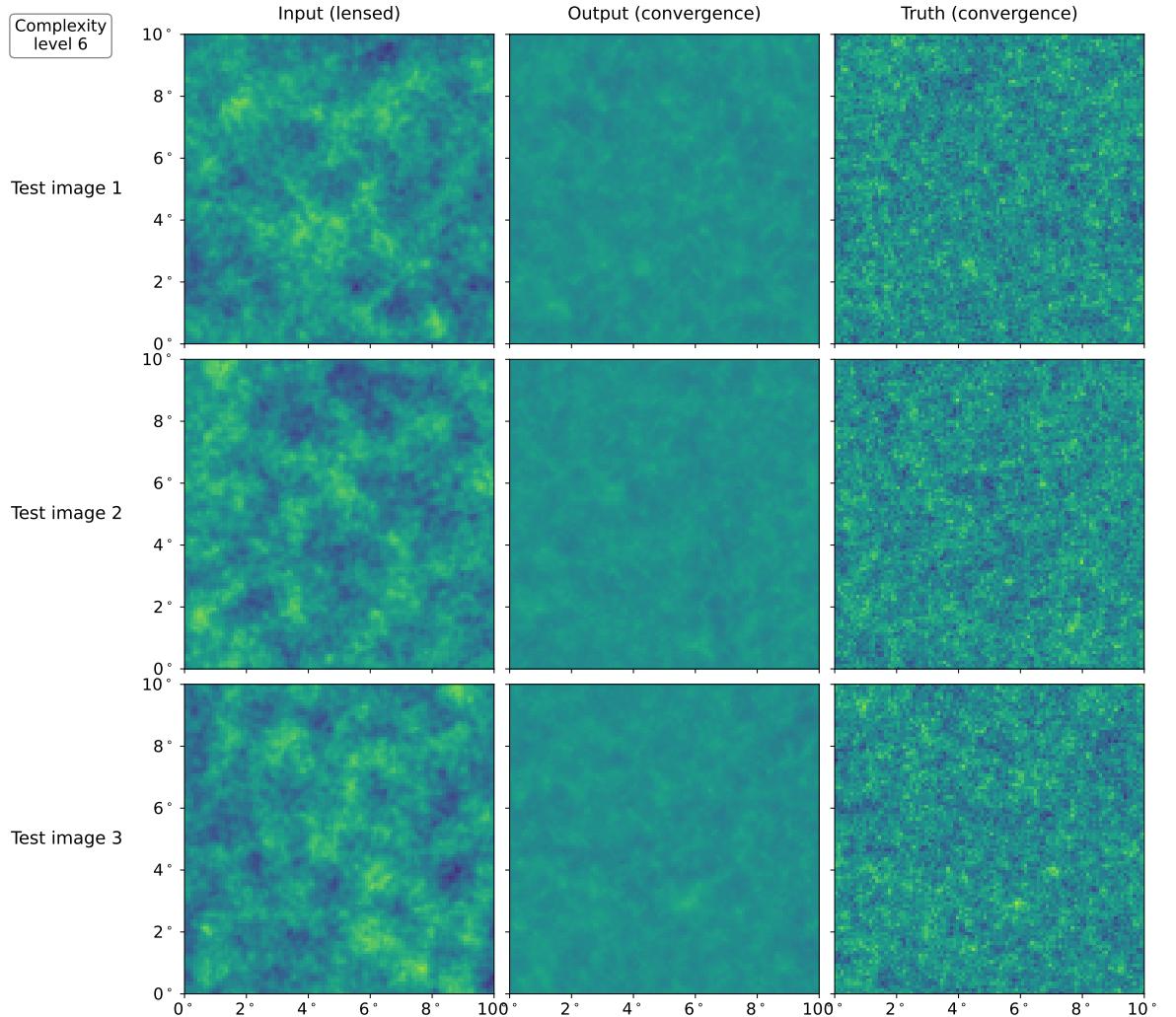


Figure 7.17. Performance of the best-performing L6_baseline model for complexity level 6 (described in Section 7.4.6) on three previously unseen test images. Rows correspond to the different test images, while the columns show, from left to right: the lensed CMB map given as input to the model, the predicted convergence map produced by the model, and the true convergence map that was used to generate the test image in the first column. The model was trained for 70 epochs on a training set of 9800 samples augmented by a factor 8, with a validation set of 200 samples which was not augmented.

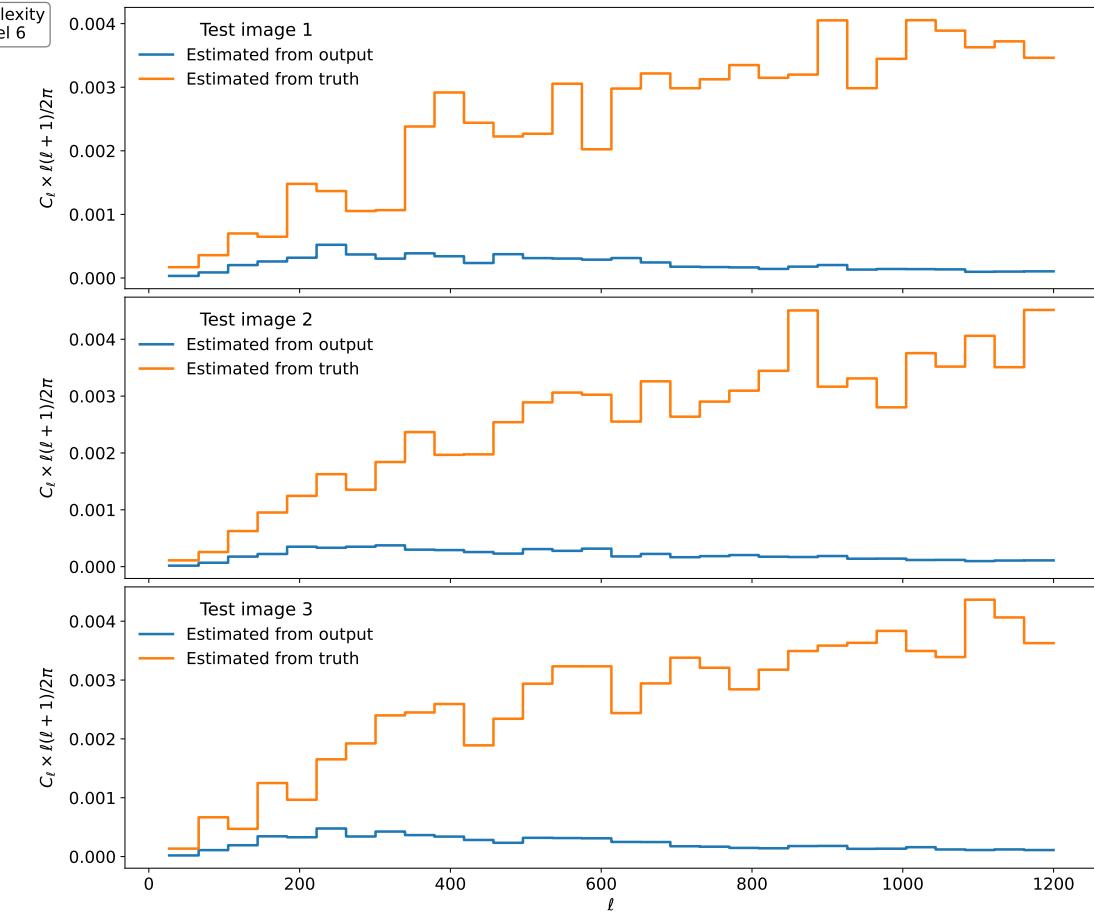


Figure 7.18. Power spectra estimated from the convergence maps shown in Figure 7.17. Rows correspond to the three different test images. The blue solid line shows the power spectrum estimated from the predicted convergence map output by the convolutional neural network, while the orange dashed line shows the power spectrum estimated from the true convergence map. Results shown are for the L5_baseline model for complexity level 6 (described in Section 7.4.6). The model was trained for 70 epochs on a training set of 9800 samples augmented by a factor 8, with a validation set of 200 samples which was not augmented.

7.5. Discussion

This chapter has presented an initial exploration of the use of convolutional neural networks for the estimation of weak lensing shear. While many issues remain outstanding, and the conditions investigated have been far short of full realism, the method has nonetheless shown promise. With a restricted signal resolution (using $\ell_{\max} = 50$ in the input convergence power spectrum) and an exaggerated lensing effect by a factor 5–30, a convolutional neural network given a lensed CMB map was able to recover the corresponding convergence map to a high degree of accuracy (Figures 7.8 and 7.10). Even with the lensing exaggeration removed, a good degree of recovery was observed (Figure 7.12). However, smaller-scale features in the convergence maps have not been recovered, and their presence seems to have hindered the recovery of the larger-scale features (Figure 7.17).

Besides imperfect performance, there have been a number of other issues. The sudden divergence of training and validation loss during training, seen most prominently in Figure 7.13, is not understood. The failure of the L5_deep model to recover from this divergence is of particular concern. This issue did not arise again, nor could records be found in the literature of similar problems. As mentioned above, it should not be permitted according to the update rules of the Adam optimiser (Equation 7.14). Another challenge is that of computational requirements, which rapidly increased as complexity was added. These arise from all steps of the data generation, model building, and training processes. Data generation is time-consuming and large training sets can require a huge volume of storage, which must be accessible to the machine(s) used for training. Training is itself a slow process and the memory requirements for an image-to-image pipeline quickly spiral as resolution and model complexity are increased, which they must be to achieve high levels of both realism and performance. In this work, these requirements at many stages forced a reduction in complexity from what could have otherwise been investigated.

Other than more computing resources, there are many ways in which performance could potentially be improved at the current level of complexity. For example, the learning rate α used in the Adam optimiser (see Section 7.2.2) has been held constant throughout this work, and while other values were explored, each was always treated as constant. Variable learning rates hold some potential promise; methods include learning rate schedulers, which vary the learning rate as a function of training epoch, or dynamic methods such as reducing the learning rate upon reaching a loss plateau. Another possibility is integrating the training data generation into the training pipeline for a limitless supply of training data, which need not be stored on disk, thereby circumventing one of the computational restrictions discussed above. This idea has not been tested, but if implemented well could potentially only slow the training process minimally. Its advantage would be to increase the rate and eventual amount of learning, which is naturally reduced each time the training set is repeated.

While remaining within the CMB lensing context, there are many additional layers of complexity that can be added. The main such angle explored here was resolution, which is still far below a realistic combination of a wide field and small pixel size. As wider fields are explored, the flat-sky approximation used here becomes more inaccurate. An extension to curved-sky geometry with appropriate pixelisation such as HEALPix is attractive but challenging, although methods for implementing convolutional neural networks in this context do exist (Perraudin et al. 2019). In addition, a telescope beam and detector noise have been neglected throughout this work, both of which must be added to simulate a realistic case. There is also the problem that a single known cosmology has been used to produce all of the training, validation, and test data. In practice, we cannot know the true cosmology, so it is necessary to take this into consideration in training and testing.

Finally, as described in the introduction to the chapter in Section 7.1, the context in which this work is motivated is not CMB lensing but instead galaxy lensing using radio observations, with an ultimate aim of developing a method to estimate weak lensing shear directly from radio visibilities. It is essential that such methods are developed and refined in advance of the influx of data predicted to arrive from the SKA in the next decade. The work presented in this chapter has not begun to even attempt to face this challenge. It should not be necessary, though, to master CMB lensing before moving onto radio galaxy lensing. It could be argued that the results presented in this chapter show sufficient promise that now is the appropriate time to move onto applying these methods towards the goal of accurate and high-precision weak lensing analyses with the SKA.

Chapter 8

Conclusions

8.1. Summary

Cosmology is in a peculiar state: on one hand, we have a model (Λ CDM, described in Chapter 1) that fits most observational data, particularly from the cosmic microwave background (CMB), with remarkable accuracy. On the other hand, almost nothing is understood about the two apparently dominant components of the Universe today—dark energy and dark matter, and it is not clear how to reconcile these components with the theories of gravity and of the standard model of particle physics.

Analysis of weak gravitational lensing by the large scale structure of the Universe has particular promise as an observational probe with which to make progress on these questions. The subtle distortions of the shapes of distant galaxies induced by the gravitational effect of large scale structure depends closely on how exactly this structure has formed and evolved as the Universe has expanded, which in turn depends on the details of dark energy, dark matter, and gravity. Precise measurements and statistical analysis of these distortions can therefore place tight constraints on the physical nature of the dominant components in the Universe.

This promise is set to be realised by an upcoming generation of weak lensing surveys, such as the *Euclid* space mission, the Rubin optical observatory, and the Square Kilometre Array radio observatory. (These surveys and others are described in Chapter 1.) This next generation will survey billions of galaxies and, for this reason, will reach unprecedented statistical precision in their cosmological constraints. This unprecedented precision will, in turn, demand unprecedented accuracy and understanding of every aspect of the analysis in order to obtain reliable results. The work presented in this thesis has helped to make progress towards this goal.

Chapters 3–6 have contributed to understanding the statistical properties of pseudo- C_ℓ estimators, which are fast two-point correlation estimators in Fourier space designed for partial-sky observations and are described in Chapter 2. These will be used for *Euclid* and have been used in a number of weak lensing analyses to date (see Chapter 2 and references therein), and yet prior to the work presented in this thesis little was known about their statistical properties. In Chapter 3, the exact joint likelihood function for an arbitrary number of pseudo- C_ℓ estimates from an arbitrary number of correlated spin-0 and spin-2 Gaussian fields was derived. While the likelihood function is not practical to evaluate in full

for a real cosmological analysis, and the Gaussian field assumption does not hold for weak lensing on all scales, this work was an essential stepping stone for the work presented in Chapter 4. This later work demonstrated, with a high degree of robustness to changes in the analysis setup or details of the underlying fields, that a Gaussian likelihood is a sufficiently accurate choice for placing constraints on dark energy with data from a *Euclid*-like survey. This means that when results obtained using a Gaussian likelihood are surprising or suspicious, we can be confident that this is not the result of an inappropriate likelihood function. A Gaussian likelihood requires a covariance matrix, and in Chapter 5 a method was developed for calculating the covariance matrix of pseudo- C_ℓ estimates, accounting for both the coupling of scales due to the effect of the mask, which describes the incomplete sky coverage, and the non-Gaussian coupling arising from non-linear structure growth on small physical scales. This covariance matrix was compared to an estimated covariance matrix measured from weak lensing simulations, with good agreement, and it was demonstrated that neglecting the non-Gaussian contributions leads to poor accuracy in parameter constraints. Finally, Chapter 6 considered the question of how to choose the number of angular bins used in a pseudo- C_ℓ analysis, which is a topic that had not previously been investigated. This question was also investigated for real-space correlation functions, and it was revealed that while the statistical precision of both estimators converges for a similar number of angular bins, the divergence of this precision for a too-broad binning is dramatically different between the two. This is an important demonstration that results derived for one estimator cannot be automatically extended to others.

Chapter 7 turned to another problem that must be addressed before reliable cosmological inference is possible with the next generation of weak lensing surveys, this time focusing on radio observations, for which shear estimation is a particular challenge. In Chapter 7, a machine learning approach to this challenge was investigated, using a simplified case of CMB lensing as a starting point. The results therein suggest that convolutional neural networks are a method of promise in this area.

8.2. Future prospects

Naturally many questions remain. For instance, while Chapter 5 provided a method for calculating the covariance matrix of pseudo- C_ℓ estimates, and presented a comparison of the result with simulations, it did not rigorously quantify the accuracy of the method at the level required for *Euclid*, nor did it compare different options at each step. How are the non-Gaussian contributions to the covariance best calculated? The super-sample covariance component dominates over the connected non-Gaussian component, and yet the latter cannot be neglected. The method used to calculate the connected non-Gaussian component in Chapter 5 is extremely slow, to the point of not being feasible for use when analysing future data from *Euclid*. An approximation was presented, but is this sufficiently

accurate? If not, can a better approximation be developed? The super-sample covariance component in Chapter 5 was calculated using the halo model. Is this approach sufficiently accurate? Should the response approach be used instead? Should the three-dimensional details of the survey volume be taken into account, describing not only the survey mask but the fact that the depth of the survey varies over its area? What about correlations between the signal and mask, which are likely to exist when dense regions are preferentially masked out? Meanwhile, the generally dominant component of the pseudo- C_ℓ covariance is the Gaussian component. In Chapter 5, this was calculated using the improved narrow kernel approximation (NKA), which was found to be significantly more accurate than the standard NKA, but is it sufficiently accurate for *Euclid*? Perhaps it is necessary to move to an approximation closer to the known exact form. And how should we properly account for the combination of cut-sky mode coupling and non-Gaussian mode coupling? This was done in an approximate manner in Chapter 5, but again it is unknown whether this is sufficiently accurate. Maybe a superior method can be derived that is still computationally feasible. Ideally answering these questions could be achieved with a large suite of fully realistic simulations, but that seems impossible when even a single realisation of the *Euclid* flagship simulations takes close to a million node hours to evaluate.

Meanwhile, Chapter 6 considered the question of how to choose the number of angular bins used for two-point estimators with *Euclid*. The results offer some useful data points but do not establish an ultimate answer to this question. On the contrary, it was found that the answer changes depending on which cosmological parameters are being constrained, how many parameters are being constrained, and other aspects such as scale cuts and noise levels. If the answer is truly so sensitive to all of these considerations then it becomes extremely complicated to be confident that the correct choice has been made for a particular analysis setup. Perhaps a universal solution can be found. On the other hand, perhaps it is best not to bin at all, at least for the power spectrum. This may be possible, but may cause problems due to the resulting length of the data vector. For the correlation function, though, this is not an option, since the data are not fundamentally discrete. Another interesting result in Chapter 6 is that the stark difference in constraining power between the power spectrum and correlation function with an extremely coarse binning can be explained by the different ways in which scales are weighed within bins between the two statistics. This raises a further question of how this weighting should be chosen, if it has the ability to make such a difference to the results. Finally, there is the accompanying question of how best to bin in redshift space, and how this may itself interact with the choice of angular binning strategy.

A further aspect of the use of any estimator that must be considered for its use for precision cosmology is the modelling of the signal, including the contribution from noise. A forward-modelling approach is being taken for the pseudo- C_ℓ estimator for *Euclid*, which means that essentially the data are left alone, and the theory prediction to the data is instead treated to contain all the contributions that have entered the observations. Modelling the signal

correctly is a big challenge: there are significant challenges on the theoretical side, but there are also many estimator effects that must be accurately modelled. How can we be confident that these cumulative effects are modelled sufficiently accurately? Comparisons between results obtained with different estimators are valuable, but it is difficult to be certain that no systematic errors remain. Different estimators naturally produce different results because of their distinct natures, such as how they weight different scales, meaning that there is no perfect agreement even in the absence of any systematic errors. This is a challenge that will require more work in order to have full confidence in the results obtained using any estimators for *Euclid*.

Moving on from considering *Euclid* and pseudo- C_ℓ estimates, Chapter 7 considered the use of convolutional neural networks (CNNs) for weak lensing estimation. The results show that CNNs for this purpose have promise, but the method developed in that chapter is far from a finished method for the ultimate aim of estimating shear from radio visibilities. The chapter focuses instead on the case of CMB lensing, but much work needs to be done in order to develop this into a method that is viable for current or future CMB experiments. Many more challenges lie ahead before the method would be suitable for radio observations with the Square Kilometre Array (SKA), which are discussed in the chapter. As with the previous chapters, there still exists a fundamental challenge with testing any candidate method, because this ideally requires full-realism simulations of SKA observations, with a sufficient number of realisations to detect subtle systematic errors, which is extremely challenging computationally.

Besides all of the questions laid out above, which follow on immediately from the work presented in this thesis, there are many more questions that must be answered in advance of receiving data from the Stage IV generation of weak lensing experiments such as *Euclid* and the SKA. Every aspect of the analysis must be intricately understood: how to treat all the data appropriately, and how to model everything theoretically. The ultimate aim is to reach a point where we can fully believe whatever the results from this upcoming generation of experiments may tell us about the nature of the Universe.

Appendix A

Proofs supplementing exact pseudo- C_ℓ likelihood

This appendix contains proofs accompanying the exact joint pseudo- C_ℓ likelihood for correlated Gaussian fields presented in Chapter 3. Section A.1 contains the derivation of the general pseudo- $a_{\ell m}$ covariance, and Section A.2 contains a proof that the joint characteristic function used in the implementation of the likelihood in Chapter 3 is identical to that presented in Good (1963).

A.1. General pseudo- $a_{\ell m}$ covariance derivation

This section contains the derivation of the general covariance matrix elements of the pseudo- $a_{\ell m}$ s.

First, the derivation of the real and imaginary parts of the general pseudo- $a_{\ell m}$ s given in Equations (3.15) and (3.16). We begin with the general form of a pseudo- $a_{\ell m}$ as a weighted sum of full-sky $a_{\ell m}$ s as given in Equation (3.13). We expand the sum over m' into separate sums for $m' < 0$ and $m' > 0$, and a term for $m' = 0$:

$$\tilde{a}_{\ell m}^{(\alpha)} = \sum_{\beta, \ell'} \left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} a_{\ell' 0}^{(\beta)} + \sum_{m' < 0} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} a_{\ell' m'}^{(\beta)} + \sum_{m' > 0} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} a_{\ell' m'}^{(\beta)} \right]. \quad (\text{A.1})$$

Writing the $m' < 0$ contributions in terms of their $m' > 0$ counterparts using Equation (3.4) gives

$$\begin{aligned} \tilde{a}_{\ell m}^{(\alpha)} &= \sum_{\beta, \ell'} \left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} a_{\ell' 0}^{(\beta)} + \sum_{m' < 0} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} (-1)^{m'} \left(a_{\ell' |m'|}^{(\beta)} \right)^* + \sum_{m' > 0} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} a_{\ell' m'}^{(\beta)} \right] \\ &= \sum_{\beta, \ell'} \left[\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} a_{\ell' 0}^{(\beta)} + \sum_{m' > 0} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} a_{\ell' m'}^{(\beta)} + (-1)^{m'} \frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \left(a_{\ell' m'}^{(\beta)} \right)^* \right) \right]. \end{aligned} \quad (\text{A.2})$$

We now expand this into real and imaginary parts as

$$\begin{aligned} \tilde{a}_{\ell m}^{(\alpha)} = & \sum_{\beta, \ell'} \left[\left[\operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} \right) + i \operatorname{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' 0}^{(\beta)}} \right) \right] \operatorname{Re} \left(a_{\ell' 0}^{(\beta)} \right) \right. \\ & + \sum_{m' > 0} \left(\left[\operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right) + i \operatorname{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' m'}^{(\beta)}} \right) \right] \left[\operatorname{Re} \left(a_{\ell' m'}^{(\beta)} \right) + i \operatorname{Im} \left(a_{\ell' m'}^{(\beta)} \right) \right] \right. \\ & \left. \left. + (-1)^{m'} \left[\operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right) + i \operatorname{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell' -m'}^{(\beta)}} \right) \right] \left[\operatorname{Re} \left(a_{\ell' m'}^{(\beta)} \right) - i \operatorname{Im} \left(a_{\ell' m'}^{(\beta)} \right) \right] \right) \right], \end{aligned} \quad (\text{A.3})$$

which, after some algebra, gives the real and imaginary parts in Equations (3.15) and (3.16).

We now proceed to the derivation of the covariance between two of the real parts, as given in Equation (3.23). The derivations of the imaginary–imaginary and real–imaginary covariance given in Equations (3.24) and (3.25) follow analogously. We begin by inserting the expression for the real part of a pseudo- $a_{\ell m}$ from Equation (3.15):

$$\begin{aligned} & \operatorname{Cov} \left(\operatorname{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right), \operatorname{Re} \left(\tilde{a}_{\ell' m'}^{(\beta)} \right) \right) \\ = & \operatorname{Cov} \left(\sum_{\gamma, \ell''} \left[\operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' 0}^{(\gamma)}} \right) \operatorname{Re} \left(a_{\ell'' 0}^{(\gamma)} \right) \right. \right. \\ & + \sum_{m'' > 0} \left(\left[\operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right) + (-1)^{m''} \operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right) \right] \operatorname{Re} \left(a_{\ell'' m''}^{(\gamma)} \right) \right. \\ & \left. \left. - \left[\operatorname{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right) - (-1)^{m''} \operatorname{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right) \right] \operatorname{Im} \left(a_{\ell'' m''}^{(\gamma)} \right) \right) \right], \\ & \sum_{\varepsilon, \ell'''} \left[\operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' 0}^{(\varepsilon)}} \right) \operatorname{Re} \left(a_{\ell''' 0}^{(\varepsilon)} \right) \right. \\ & + \sum_{m''' > 0} \left(\left[\operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' m'''}^{(\varepsilon)}} \right) + (-1)^{m'''} \operatorname{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' -m'''}^{(\varepsilon)}} \right) \right] \operatorname{Re} \left(a_{\ell''' m'''}^{(\varepsilon)} \right) \right. \\ & \left. \left. - \left[\operatorname{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' m'''}^{(\varepsilon)}} \right) - (-1)^{m'''} \operatorname{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' -m'''}^{(\varepsilon)}} \right) \right] \operatorname{Im} \left(a_{\ell''' m'''}^{(\varepsilon)} \right) \right) \right). \end{aligned} \quad (\text{A.4})$$

We expand this into a linear sum of full-sky $a_{\ell m}$ covariances, noting that the terms involving covariance between real and imaginary full-sky $a_{\ell m}$ s immediately vanish, as do terms involving covariance between $m = 0$ and $m > 0$ $a_{\ell m}$ s:

$$\begin{aligned}
& \text{Cov} \left(\text{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right), \text{Re} \left(\tilde{a}_{\ell' m'}^{(\beta)} \right) \right) \\
&= \sum_{\gamma, \varepsilon} \sum_{\ell'' \ell''' m''} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' 0}^{(\gamma)}} \right) \text{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' 0}^{(\varepsilon)}} \right) \text{Cov} \left(\text{Re} \left(a_{\ell'' 0}^{(\gamma)} \right), \text{Re} \left(a_{\ell''' 0}^{(\varepsilon)} \right) \right) \right. \\
&\quad + \sum_{\substack{m'' > 0 \\ m''' > 0}} \left(\left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right) + (-1)^{m''} \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right) \right] \right. \\
&\quad \times \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' m'''}^{(\varepsilon)}} \right) + (-1)^{m'''} \text{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' -m'''}^{(\varepsilon)}} \right) \right] \\
&\quad \times \text{Cov} \left(\text{Re} \left(a_{\ell'' m''}^{(\gamma)} \right), \text{Re} \left(a_{\ell''' m'''}^{(\varepsilon)} \right) \right) \\
&\quad + \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right) - (-1)^{m''} \text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right) \right] \\
&\quad \times \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' m'''}^{(\varepsilon)}} \right) - (-1)^{m'''} \text{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell''' -m'''}^{(\varepsilon)}} \right) \right] \\
&\quad \times \text{Cov} \left(\text{Im} \left(a_{\ell'' m''}^{(\gamma)} \right), \text{Im} \left(a_{\ell''' m'''}^{(\varepsilon)} \right) \right) \left. \right] . \tag{A.5}
\end{aligned}$$

The full-sky covariances are given by

$$\text{Cov} \left(\text{Re} \left(a_{\ell m}^{(\alpha)} \right), \text{Re} \left(a_{\ell' m'}^{(\beta)} \right) \right) = \begin{cases} C_{\ell}^{\alpha\beta} \delta_{\ell\ell'} \delta_{mm'} & m = 0; \\ C_{\ell}^{\alpha\beta} \delta_{\ell\ell'} \delta_{mm'}/2 & m > 0, \end{cases} \tag{A.6}$$

and for $m > 0$,

$$\text{Cov} \left(\text{Im} \left(a_{\ell m}^{(\alpha)} \right), \text{Im} \left(a_{\ell' m'}^{(\beta)} \right) \right) = C_{\ell}^{\alpha\beta} \delta_{\ell\ell'} \delta_{mm'}/2. \tag{A.7}$$

Inserting these into Equation (A.5), evaluating the delta functions and rearranging, we obtain

$$\begin{aligned}
& \text{Cov} \left(\text{Re} \left(\tilde{a}_{\ell m}^{(\alpha)} \right), \text{Re} \left(\tilde{a}_{\ell' m'}^{(\beta)} \right) \right) \\
&= \sum_{\gamma, \varepsilon} \sum_{\ell''} C_{\ell''}^{\gamma \varepsilon} \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' 0}^{(\gamma)}} \right) \text{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' 0}^{(\varepsilon)}} \right) \right. \\
&\quad + \frac{1}{2} \sum_{m'' > 0} \left(\left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right) + (-1)^{m''} \text{Re} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right) \right] \right. \\
&\quad \times \left[\text{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) + (-1)^{m''} \text{Re} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) \right] \\
&\quad + \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' m''}^{(\gamma)}} \right) - (-1)^{m''} \text{Im} \left(\frac{\partial \tilde{a}_{\ell m}^{(\alpha)}}{\partial a_{\ell'' -m''}^{(\gamma)}} \right) \right] \\
&\quad \times \left. \left. \left[\text{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' m''}^{(\varepsilon)}} \right) - (-1)^{m''} \text{Im} \left(\frac{\partial \tilde{a}_{\ell' m'}^{(\beta)}}{\partial a_{\ell'' -m''}^{(\varepsilon)}} \right) \right] \right] \right]. \tag{A.8}
\end{aligned}$$

Finally, we use the identities that for complex A and B ,

$$\text{Re } A \text{Re } B + \text{Im } A \text{Im } B = \text{Re} (A^* B); \tag{A.9}$$

$$\text{Re } A \text{Re } B - \text{Im } A \text{Im } B = \text{Re} (AB) \tag{A.10}$$

to obtain the final covariance given in Equation (3.23). The variance of a single real part is a special case of Equation (3.23) having $\alpha = \beta$, $\ell = \ell'$ and $m = m'$.

A.2. Equivalence of the two forms of the joint characteristic function

This section contains a proof that the expression for the joint characteristic function used in the implementation of the likelihood used in Chapter 3, given in Equation (3.48) and denoted here by φ_{alt} , is mathematically equivalent to the known analytic form from Good (1963), given in Equation (3.33) and denoted here by φ_{Good} .

The form given in Good (1963) is written in terms of the determinant of the complex square matrix which we will denote by \mathbf{X} :

$$\mathbf{X} = \mathbf{I} - 2i \sum_{\ell} \sum_{\alpha \beta} t_{\ell}^{\alpha \beta} \mathbf{M}_{\ell}^{\alpha \beta} \Sigma, \tag{A.11}$$

where \mathbf{I} is the identity matrix. The determinant of any matrix is equal to the product of its eigenvalues, so we may write φ_{Good} as

$$\varphi_{\text{Good}} (\mathbf{t}) = \prod_j \lambda_j^{-1/2}, \quad \lambda_j \in \lambda (\mathbf{X}). \tag{A.12}$$

The eigenvalues of \mathbf{X} , $\lambda(\mathbf{X})$, are defined by the eigenvalue equation,

$$|\mathbf{X} - \lambda_j \mathbf{I}| = 0 \quad \forall \lambda_j \in \lambda(\mathbf{X}). \quad (\text{A.13})$$

Now let us define another matrix \mathbf{Y} such that $\mathbf{X} = \mathbf{I} - \mathbf{Y}$. Inserting this into Equation (A.13), we obtain

$$\left| -[\mathbf{Y} - (1 - \lambda_j) \mathbf{I}] \right| = 0 \quad \forall \lambda_j \in \lambda(\mathbf{X}). \quad (\text{A.14})$$

For any matrix \mathbf{A} , the determinant of $-\mathbf{A}$ is given by

$$|-\mathbf{A}| = \pm |\mathbf{A}|, \quad (\text{A.15})$$

depending on whether the rank of \mathbf{A} is even or odd. Using this fact in Equation (A.14) implies that

$$\left| \mathbf{Y} - (1 - \lambda_j) \mathbf{I} \right| = 0 \quad \forall \lambda_j \in \lambda(\mathbf{X}). \quad (\text{A.16})$$

Therefore, the eigenvalues of \mathbf{X} and \mathbf{Y} are related as

$$\lambda(\mathbf{Y}) = 1 - \lambda(\mathbf{X}), \quad \text{and hence} \quad \lambda(\mathbf{X}) = 1 - \lambda(\mathbf{Y}). \quad (\text{A.17})$$

This allows us to write φ_{Good} in terms of the eigenvalues of \mathbf{Y} ,

$$\varphi_{\text{Good}}(\mathbf{t}) = \prod_j (1 - \lambda_j)^{-1/2}, \quad \lambda_j \in \lambda(\mathbf{Y}). \quad (\text{A.18})$$

Finally, we use the fact that the eigenvalues of a scalar multiple of a matrix are equal to the scalar multiplied by the original matrix,

$$\lambda(\alpha \mathbf{A}) = \alpha \lambda(\mathbf{A}), \quad (\text{A.19})$$

to extract the multiple of $2i$ from \mathbf{Y} :

$$\lambda(\mathbf{Y}) = \lambda \left(2i \sum_{\ell} \sum_{\alpha\beta} t_{\ell}^{\alpha\beta} \mathbf{M}_{\ell}^{\alpha\beta} \boldsymbol{\Sigma} \right) = 2i \times \lambda \left(\sum_{\ell} \sum_{\alpha\beta} t_{\ell}^{\alpha\beta} \mathbf{M}_{\ell}^{\alpha\beta} \boldsymbol{\Sigma} \right), \quad (\text{A.20})$$

which we insert into Equation (A.18) to obtain the alternative form,

$$\begin{aligned} \varphi_{\text{Good}}(\mathbf{t}) &= \prod_j (1 - 2i\lambda_j)^{-1/2}, \quad \lambda_j \in \lambda \left(\sum_{\ell} \sum_{\alpha\beta} t_{\ell}^{\alpha\beta} \mathbf{M}_{\ell}^{\alpha\beta} \boldsymbol{\Sigma} \right) \\ &= \varphi_{\text{alt}}(\mathbf{t}). \end{aligned} \quad (\text{A.21})$$

Bibliography

- Aaboud, M., Aad, G., Abbott, B., Abdinov, O., Abeloos, B., Abidi, S., AbouZeid, O., Abraham, N., Abramowicz, H., Abreu, H. et al. 2018,
European Physical Journal C 78(12), 995.
- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M. et al. 2016,
preprint (arXiv:1603.04467).
- Abazajian, K. N., Adshead, P., Ahmed, Z., Allen, S. W., Alonso, D., Arnold, K. S., Baccigalupi, C., Bartlett, J. G., Battaglia, N., Benson, B. A. et al. 2016,
preprint (arXiv:1610.02743).
- Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X. et al. 2016,
Physical Review Letters 116(6), 61102.
- Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., Adya, V. B. et al. 2017,
Nature 551(7678), 85.
- Abbott, T. M., Abdalla, F. B., Alarcon, A., Aleksić, J., Allam, S., Allen, S., Amara, A., Annis, J., Asorey, J., Avila, S. et al. 2018,
Physical Review D 98(4), 1.
- Abramovici, A., Althouse, W. E., Drever, R. W. P., Gürsel, Y., Kawamura, S., Raab, F. J., Shoemaker, D., Sievers, L., Spero, R. E., Thorne, K. S. et al. 1992,
Science 256(5055), 325.
- Accadia, T., Acernese, F., Alshourbagy, M., Amico, P., Antonucci, F., Aoudia, S., Arnaud, N., Arnault, C., Arun, K. G., Astone, P. et al. 2012,
Journal of Instrumentation 7(03), 03012.
- Ade, P., Aguirre, J., Ahmed, Z., Aiola, S., Ali, A., Alonso, D., Alvarez, M. A., Arnold, K., Ashton, P., Austermann, J. et al. 2019,
Journal of Cosmology and Astroparticle Physics 2019(02), 056.
- Adler, R. J., Casey, B. & Jacob, O. C. 1995,
American Journal of Physics 63(7), 620.
- Aihara, H., Arimoto, N., Armstrong, R., Arnouts, S., Bahcall, N. A., Bickerton, S., Bosch, J., Bundy, K., Capak, P. L., Chan, J. H. H. et al. 2018,
Publications of the Astronomical Society of Japan 70(SP1), S4.
- Aihara, H., AlSayyad, Y., Ando, M., Armstrong, R., Bosch, J., Egami, E., Furusawa, H., Furusawa, J., Harasawa, S., Harikane, Y. et al. 2022,
Publications of the Astronomical Society of the Pacific 74(2), 247.
- Akutsu, T., Ando, M., Arai, K., Arai, Y., Araki, S., Araya, A., Aritomi, N., Asada, H., Aso, Y., Atsuta, S. et al. 2019,

- Nature Astronomy 3(1), 35.
- Albrecht, A., Bernstein, G., Cahn, R., Freedman, W. L., Hewitt, J., Hu, W., Huth, J., Kamionkowski, M., Kolb, E. W., Knox, L. et al. 2006,
- Alonso, D., Sanchez, J. & Slosar, A. 2019,
- Monthly Notices of the Royal Astronomical Society 484(3), 4127.
- Alpher, R. A. & Herman, R. C. 1948,
- Physical Review 74(12), 1737.
- Alpher, R. A. & Herman, R. C. 1948,
- Nature 162(4124), 774.
- Alsing, J. & Wandelt, B. 2019,
- Monthly Notices of the Royal Astronomical Society 488(4), 5093.
- Amaro-Seoane, P., Audley, H., Babak, S., Baker, J., Barausse, E., Bender, P., Berti, E., Binetruy, P., Born, M., Bortoluzzi, D. et al. 2017,
- preprint (arXiv:1702.00786).
- Amaro, V., Cavaudi, S., Brescia, M., Vellucci, C., Longo, G., Bilicki, M., de Jong, J. T. A., Tortora, C., Radovich, M., Napolitano, N. R. et al. 2019,
- Monthly Notices of the Royal Astronomical Society 482(3), 3116.
- Amon, A., Gruen, D., Troxel, M. A., MacCrann, N., Dodelson, S., Choi, A., Doux, C., Secco, L., Samuroff, S., Krause, E. et al. 2022,
- Physical Review D 105(2), 023514.
- Anand, G. S., Tully, R. B., Rizzi, L., Riess, A. G. & Yuan, W. 2022,
- The Astrophysical Journal 932(1), 15.
- Anderson, L., Aubourg, E., Bailey, S., Bizyaev, D., Blanton, M., Bolton, A. S., Brinkmann, J., Brownstein, J. R., Burden, A., Cuesta, A. J. et al. 2012,
- Monthly Notices of the Royal Astronomical Society 427(4), 3435.
- Anderson, L., Aubourg, É., Bailey, S., Beutler, F., Bhardwaj, V., Blanton, M., Bolton, A. S., Brinkmann, J., Brownstein, J. R., Burden, A. et al. 2014,
- Monthly Notices of the Royal Astronomical Society 441(1), 24.
- Andrade, K. E., Fuson, J., Gad-Nasr, S., Kong, D., Minor, Q., Roberts, M. G. & Kaplinghat, M. 2022,
- Monthly Notices of the Royal Astronomical Society 510(1), 54.
- Aniyan, A. K. & Thorat, K. 2017,
- The Astrophysical Journal Supplement Series 230(2), 20.
- Arai, S. & Nishizawa, A. 2018,
- Physical Review D 97(10), 104038.
- Asgari, M., Heymans, C., Hildebrandt, H., Miller, L., Schneider, P., Amon, A., Choi, A., Erben, T., Georgiou, C., Harnois-Deraps, J. et al. 2019,
- Astronomy & Astrophysics 624, A134.
- Asgari, M., Lin, C.-A., Joachimi, B., Giblin, B., Heymans, C., Hildebrandt, H., Kannawadi, A., Stölzner, B., Tröster, T., van den Busch, J. L. et al. 2021,
- Astronomy & Astrophysics 645, A104.

- ATLAS Collaboration 2019,
 preprint (arXiv:1911.12606).
- Bacon, D. J., Refregier, A. R. & Ellis, R. S. 2000,
Monthly Notices of the Royal Astronomical Society 318(2), 625.
- Baldi, P., Bian, J., Hertel, L. & Li, L. 2019,
Physical Review D 99(1), 12011.
- Bartolo, N., Caprini, C., Domcke, V., Figueroa, D. G., Garcia-Bellido, J., Guzzetti, M. C.,
 Liguori, M., Matarrese, S., Peloso, M., Petiteau, A. et al. 2016,
Journal of Cosmology and Astroparticle Physics 2016(12), 26.
- Barreira, A. & Schmidt, F. 2017,
Journal of Cosmology and Astroparticle Physics 2017(6), 53.
- Barreira, A. & Schmidt, F. 2017,
Journal of Cosmology and Astroparticle Physics 2017(11), 51.
- Barreira, A., Krause, E. & Schmidt, F. 2018,
Journal of Cosmology and Astroparticle Physics 2018(06), 015.
- Barreira, A., Krause, E. & Schmidt, F. 2018,
Journal of Cosmology and Astroparticle Physics 2018(10), 053.
- Bardeen, J. M., Steinhardt, P. J. & Turner, M. S. 1983,
Physical Review D 28(4), 679.
- Battye, R. A., Pace, F. & Trinh, D. 2018,
Physical Review D 98(2), 23504.
- Baumann, D. & McAllister, L. 2015,
Inflation and String Theory, Cambridge University Press.
- Beltrán Jiménez, J., Piazza, F. & Velten, H. 2016,
Physical Review Letters 116(6), 61101.
- Benabed, K., Cardoso, J.-F., Prunet, S. & Hivon, E. 2009,
Monthly Notices of the Royal Astronomical Society 400(1), 219.
- Benjamin, J., Van Waerbeke, L., Heymans, C., Kilbinger, M., Erben, T., Hildebrandt, H.,
 Hoekstra, H., Kitching, T. D., Mellier, Y., Miller, L. et al. 2013,
Monthly Notices of the Royal Astronomical Society 431(2), 1547.
- Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., Smith,
 K. M., Hill, R. S., Gold, B., Halpern, M. et al. 2013,
The Astrophysical Journal Supplement Series 208(2), 20.
- Bennett, C. L., Banday, A. J., Górski, K. M., Hinshaw, G., Jackson, P., Keegstra, P., Kogut, A.,
 Smoot, G. F., Wilkinson, D. T. & Wright, E. L. 1996,
The Astrophysical Journal 464(1), L1.
- Beutler, F., Blake, C., Colless, M., Jones, D. H., Staveley-Smith, L., Campbell, L., Parker, Q.,
 Saunders, W. & Watson, F. 2011,
Monthly Notices of the Royal Astronomical Society 416(4), 3017.
- Beutler, F., Saito, S., Seo, H.-J., Brinkmann, J., Dawson, K. S., Eisenstein, D. J., Font-Ribera,
 A., Ho, S., McBride, C. K., Montesano, F. et al. 2014,

- Monthly Notices of the Royal Astronomical Society 443(2), 1065.
- Bilicki, M., Hoekstra, H., Brown, M. J. I., Amaro, V., Blake, C., Cavuoti, S., de Jong, J. T. A., Georgiou, C., Hildebrandt, H., Wolf, C. et al. 2018,
Astronomy & Astrophysics 616, A69.
- Blake, C., Kazin, E. A., Beutler, F., Davis, T. M., Parkinson, D., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S. et al. 2011,
Monthly Notices of the Royal Astronomical Society 418(3), 1707.
- Blazek, J., Vlah, Z. & Seljak, U. 2015,
Journal of Cosmology and Astroparticle Physics 2015(08), 15.
- Blake, C., Joudaki, S., Heymans, C., Choi, A., Erben, T., Harnois-Deraps, J., Hildebrandt, H., Joachimi, B., Nakajima, R., van Waerbeke, L. et al. 2016,
Monthly Notices of the Royal Astronomical Society 456(3), 2806.
- Blas, D., Ivanov, M. M., Sawicki, I. & Sibiryakov, S. 2016,
Soviet Journal of Experimental and Theoretical Physics Letters 103(10), 624.
- Blanton, M. R., Bershadsky, M. A., Abolfathi, B., Albareti, F. D., Allende Prieto, C., Almeida, A., Alonso-García, J., Anders, F., Anderson, S. F., Andrews, B. et al. 2017,
The Astronomical Journal 154(1), 28.
- Blazek, J. A., MacCrann, N., Troxel, M. A. & Fang, X. 2019,
Physical Review D 100(10), 103506.
- Blot, L., Corasaniti, P. S., Amendola, L. & Kitching, T. D. 2016,
Monthly Notices of the Royal Astronomical Society 458(4), 4462.
- Blomqvist, M., du Mas des Bourboux, H., Busca, N. G., de Sainte Agathe, V., Rich, J., Balland, C., Bautista, J. E., Dawson, K., Font-Ribera, A., Guy, J. et al. 2019,
Astronomy & Astrophysics 629, A86.
- Blumenthal, G. R., Pagels, H. & Primack, J. R. 1982,
Nature 299(5878), 37.
- Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. 1984,
Nature 311, 517.
- Boksenberg, A. 1985,
Vistas in Astronomy 28, 531.
- Bonvin, V., Courbin, F., Suyu, S. H., Marshall, P. J., Rusu, C. E., Sluse, D., Tewes, M., Wong, K. C., Collett, T., Fassnacht, C. D. et al. 2017,
Monthly Notices of the Royal Astronomical Society 465(4), 4914.
- Bond, J. R., Szalay, A. S. & Turner, M. S. 1982,
Physical Review Letters 48(23), 1636.
- Bond, J. R., Jaffe, A. H. & Knox, L. 1998,
Physical Review D 57(4), 2117.
- Bonnett, C. 2015,
Monthly Notices of the Royal Astronomical Society 449(1), 1043.
- Bose, B., Koyama, K. & Winther, H. A. 2019,
Journal of Cosmology and Astroparticle Physics 2019(10), 21.

- Bose, B., Pourtsidou, A., Marković, K. & Beutler, F. 2020,
Monthly Notices of the Royal Astronomical Society 493(4), 5301.
- Boucaud, A., Huertas-Company, M., Heneka, C., Ishida, E. E. O., Sedaghat, N., de Souza, R. S., Moews, B., Dole, H., Castellano, M., Merlin, E. et al. 2020,
Monthly Notices of the Royal Astronomical Society 491(2), 2481.
- Boyle, M., Stein, L. C. & Gross, J. 2021,
https://github.com/moble/spherical_functions.
- Boyle, M. et al. 2022,
<https://github.com/moble/quaternion>.
- Bridle, S. & King, L. 2007,
New Journal of Physics 9(12), 444.
- Brown, M. L., Taylor, A. N., Hambly, N. C. & Dye, S. 2002,
Monthly Notices of the Royal Astronomical Society 333(3), 501.
- Brown, M. L., Taylor, A. N., Bacon, D. J., Gray, M. E., Dye, S., Meisenheimer, K. & Wolf, C. 2003,
Monthly Notices of the Royal Astronomical Society 341(1), 100.
- Brown, M. L., Castro, P. G. & Taylor, A. N. 2005,
Monthly Notices of the Royal Astronomical Society 360(4), 1262.
- Brown, M. L. & Battye, R. A. 2011,
Monthly Notices of the Royal Astronomical Society 410(3), 2057.
- Brown, M. L., Bacon, D. J., Camera, S., Harrison, I., Joachimi, B., Metcalf, R. B., Pourtsidou, A., Takahashi, K., Zuntz, J. A., Abdalla, F. B. et al. 2015,
Proceedings of Advancing Astrophysics with the Square Kilometre Array.
- Bull, P., Ferreira, P. G., Patel, P. & Santos, M. G. 2015,
The Astrophysical Journal 803(1), 21.
- Burns, C. R., Parent, E., Phillips, M. M., Stritzinger, M., Krisciunas, K., Suntzeff, N. B., Hsiao, E. Y., Contreras, C., Anais, J., Boldt, L. et al. 2018,
The Astrophysical Journal 869(1), 56.
- Cabré, A., Fosalba, P., Gaztañaga, E. & Manera, M. 2007,
Monthly Notices of the Royal Astronomical Society 381(4), 1347.
- Cai, R.-G., Pi, S., Wang, S.-J. & Yang, X.-Y. 2019,
Journal of Cosmology and Astroparticle Physics 2019(10), 59.
- Caldeira, J., Wu, W. L. K., Nord, B., Avestruz, C., Trivedi, S. & Story, K. T. 2019,
Astronomy and Computing 28, 100307.
- Camera, S., Harrison, I., Bonaldi, A. & Brown, M. L. 2017,
Monthly Notices of the Royal Astronomical Society 464(4), 4747.
- Camacho, H., Kokron, N., Andrade-Oliveira, F., Rosenfeld, R., Lima, M., Lacasa, F., Sobreira, F., da Costa, L. N., Avila, S., Chan, K. C. et al. 2019,
Monthly Notices of the Royal Astronomical Society 487(3), 3870.
- Camacho, H., Andrade-Oliveira, F., Troja, A., Rosenfeld, R., Faga, L., Gomes, R., Doux, C., Fang, X., Lima, M., Miranda, V. et al. 2021,

- preprint (arXiv:2111.07203).
- Caprini, C., Figueroa, D. G., Flauger, R., Nardini, G., Peloso, M., Pieroni, M., Ricciardone, A. & Tasinato, G. 2019,
Journal of Cosmology and Astroparticle Physics 2019(11), 17.
- Carlstrom, J. E., Ade, P. A. R., Aird, K. A., Benson, B. A., Bleem, L. E., Busetti, S., Chang, C. L., Chauvin, E., Cho, H.-M., Crawford, T. M. et al. 2011,
Publications of the Astronomical Society of the Pacific 123(903), 568.
- Carron, J. 2013,
Astronomy & Astrophysics 551, A88.
- Casarini, L., Vacca, G. L., Amendola, L., Bonometto, S. A. & Macciò, A. V. 2011,
Journal of Cosmology and Astroparticle Physics 2011(03), 26.
- Challinor, A. & Chon, G. 2005,
Monthly Notices of the Royal Astronomical Society 360(2), 509.
- Chang, C., Marshall, P. J., Jernigan, J. G., Peterson, J. R., Kahn, S. M., Gull, S. F., AlSayyad, Y., Ahmad, Z., Bankert, J., Bard, D. et al. 2012,
Monthly Notices of the Royal Astronomical Society 427(3), 2572.
- Charnock, T. & Moss, A. 2017,
The Astrophysical Journal 837(2), L28.
- Chen, Y., Jiang, H., Li, C., Jia, X. & Ghamisi, P. 2016,
IEEE Transactions on Geoscience and Remote Sensing 54(10), 6232.
- Chiang, C.-T., Wagner, C., Schmidt, F. & Komatsu, E. 2014,
Journal of Cosmology and Astroparticle Physics 2014(05), 48.
- Chisari, N. E., Alonso, D., Krause, E., Leonard, C. D., Bull, P., Neveu, J., Villarreal, A., Singh, S., McClintock, T., Ellison, J. et al. 2019,
The Astrophysical Journal Supplement Series 242(1), 2.
- Chluba, J., Kogut, A., Patil, S. P., Abitbol, M. H., Aghanim, N., Ali-Haimoud, Y., Amin, M. A., Aumont, J., Bartolo, N., Basu, K. et al. 2019,
Bulletin of the American Astronomical Society 51(3).
- Chluba, J., Abitbol, M. H., Aghanim, N., Ali-Haimoud, Y., Alvarez, M., Basu, K., Bolliet, B., Burigana, C., de Bernardis, P., Delabrouille, J. et al. 2021,
Experimental Astronomy 51, 1515.
- Chon, G., Challinor, A., Prunet, S., Hivon, E. & Szapudi, I. 2004,
Monthly Notices of the Royal Astronomical Society 350(3), 914.
- Chollet, F. et al. 2015,
<https://github.com/fchollet/keras>.
- Cimatti, A., Roberto, M., Baugh, C., Beckwith, S. V. W., Content, R., Daddi, E., De Lucia, G., Garilli, B., Guzzo, L., Kauffmann, G. et al. 2009,
Experimental Astronomy 23(1), 39.
- Clerkin, L., Kirk, D., Manera, M., Lahav, O., Abdalla, F., Amara, A., Bacon, D., Chang, C., Gaztañaga, E., Hawken, A. et al. 2016,
Monthly Notices of the Royal Astronomical Society 466(2), 1444.

- Clifton, T., Ferreira, P. G., Padilla, A. & Skordis, C. 2012,
Physics Reports 513(1), 1.
- Clowe, D., Gonzalez, A. & Markevitch, M. 2004,
The Astrophysical Journal 604(2), 596.
- Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C. & Zaritsky, D. 2006,
The Astrophysical Journal 648(2), L109.
- Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R. et al. 2005,
Monthly Notices of the Royal Astronomical Society 362(2), 505.
- Cooray, A. & Hu, W. 2001,
The Astrophysical Journal 554(1), 56.
- Cooray, A. & Sheth, R. 2002,
Physics Reports 372(1), 1.
- Copeland, D., Taylor, A. & Hall, A. 2018,
Monthly Notices of the Royal Astronomical Society 480(2), 2247.
- Cordero, J. P., Harrison, I., Rollins, R. P., Bernstein, G., Bridle, S., Alarcon, A., Alves, O., Amon, A., Andrade-Oliveira, F., Camacho, H. et al. 2022,
Monthly Notices of the Royal Astronomical Society 511(2), 2170.
- Coulton, W. R., Armstrong, R., Smith, K. M., Lupton, R. H. & Spergel, D. N. 2018,
The Astronomical Journal 155(6), 258.
- Dark Energy Survey Collaboration 2005,
preprint (arXiv:astro-ph/0510346).
- Daubechies, I., DeVore, R., Foucart, S., Hanin, B. & Petrova, G. 2019,
preprint (arXiv:1905.02199).
- Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. 1985,
The Astrophysical Journal 292, 371.
- Dawson, W. A., Schneider, M. D., Tyson, J. A. & Jee, M. J. 2016,
The Astrophysical Journal 816(1), 11.
- de Sainte Agathe, V., Balland, C., du Mas des Bourboux, H., Busca, N. G., Blomqvist, M., Guy, J., Rich, J., Font-Ribera, A., Pieri, M. M., Bautista, J. E. et al. 2019,
Astronomy & Astrophysics 629, A85.
- Delubac, T., Bautista, J. E., Busca, N. G., Rich, J., Kirkby, D., Bailey, S., Font-Ribera, A., Slosar, A., Lee, K.-G., Pieri, M. M. et al. 2015,
Astronomy & Astrophysics 574, A59.
- Deshpande, A. C., Taylor, P. L. & Kitching, T. D. 2020,
Physical Review D 102(8), 83535.
- DES Collaboration, Abbott, T., Aguena, M., Alarcon, A., Allam, S., Alves, O., Amon, A., Andrade-Oliveira, F., Annis, J., Avila, S. et al. 2022,
Physical Review D 105(2), 023520.
- Dhawan, S., Jha, S. W. & Leibundgut, B. 2018,

- Astronomy & Astrophysics 609, A72.
- Di Valentino, E., Melchiorri, A. & Silk, J. 2015,
 Physical Review D 92(12), 121302.
- Di Valentino, E., Mena, O., Pan, S., Visinelli, L., Yang, W., Melchiorri, A., Mota, D. F., Riess, A. G. & Silk, J. 2021,
 Classical and Quantum Gravity 38(15), 153001.
- Diaz Rivero, A. & Dvorkin, C. 2020,
 Physical Review D 102(10), 103507.
- Dicke, R. H., Peebles, P. E., Roll, P. G. & Wilkinson, D. T. 1965,
 The Astrophysical Journal 142, 414.
- Dieleman, S., Willett, K. W. & Dambare, J. 2015,
 Monthly Notices of the Royal Astronomical Society 450(2), 1441.
- D'Isanto, A. & Polsterer, K. L. 2018,
 Astronomy & Astrophysics 609, A111.
- De Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., Valentijn, E. A., Consortiums, K. & Astro-WISE 2013,
 Experimental Astronomy 35(1-2), 25.
- Doroshkevich, A. G. & Novikov, I. D. 1964,
 Soviet Physics-Doklady 9(2), 111.
- Driver, S. P. & Robotham, A. S. G. 2010,
 Monthly Notices of the Royal Astronomical Society 407(4), 2131.
- Dyson, F. W., Eddington, A. S. & Davidson, C. 1920,
 Philosophical Transactions of the Royal Society of London Series A 220, 291.
- Dyson, F. W. 1917,
 Monthly Notices of the Royal Astronomical Society 77, 445.
- Efstathiou, G. 2004,
 Monthly Notices of the Royal Astronomical Society 349(2), 603.
- Eifler, T., Krause, E., Dodelson, S., Zentner, A. R., Hearin, A. P. & Gnedin, N. Y. 2015,
 Monthly Notices of the Royal Astronomical Society 454(3), 2451.
- Einstein, A. 1916,
 Annalen der Physik 354(7), 769.
- Einstein, A. 1917,
 Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, 142.
- Einstein, A. 1918,
 Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, 154.
- Einstein, A. 1936,
 Science 84(2188), 506.
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z. et al. 2005,
 The Astrophysical Journal 633(2), 560.
- Elsner, F. & Wandelt, B. D. 2012,

- Astronomy & Astrophysics 542, A60.
- Eriksen, M. & Hoekstra, H. 2018,
Monthly Notices of the Royal Astronomical Society 477(3), 3433.
- Erler, J. & Schott, M. 2019,
Progress in Particle and Nuclear Physics 106, 68.
- Euclid Collaboration: Blanchard, A., Camera, S., Carbone, C., Cardone, V. F., Casas, S., Clesse, S., Ilić, S., Kilbinger, M., Kitching, T., Kunz, M. et al. 2020,
Astronomy & Astrophysics 642, A191.
- Euclid Collaboration: Pocino, A., Tutusaus, I., Castander, F. J., Fosalba, P., Crocce, M., Porredon, A., Camera, S., Cardone, V., Casas, S., Kitching, T. et al. 2021,
Astronomy & Astrophysics 655, A44.
- Euclid Collaboration: Scaramella, R., Amiaux, J., Mellier, Y., Burigana, C., Carvalho, C., Cuillandre, J.-C., Da Silva, A., Derosa, A., Dinis, J., Maiorano, E. et al. 2022,
Astronomy & Astrophysics 662, A112.
- Fabbian, G., Carron, J., Lewis, A. & Lembo, M. 2021,
Physical Review D 103(4), 43535.
- Fang, X., Eifler, T. & Krause, E. 2020,
Monthly Notices of the Royal Astronomical Society 497(3), 2699.
- Fang, X., Krause, E., Eifler, T. & MacCrann, N. 2020,
Journal of Cosmology and Astroparticle Physics 2020(05), 010.
- Fixsen, D. J., Cheng, E. S., Gales, J. M., Mather, J. C., Shafer, R. A. & Wright, E. L. 1996,
The Astrophysical Journal 473, 576.
- Flaugher, B., Diehl, H. T., Honscheid, K., Abbott, T. M. C., Alvarez, O., Angstadt, R., Annis, J. T., Antonik, M., Ballester, O., Beaufort, L. et al. 2015,
The Astronomical Journal 150(5), 150.
- Fortuna, M. C., Hoekstra, H., Joachimi, B., Johnston, H., Chisari, N. E., Georgiou, C. & Mahony, C. 2021,
Monthly Notices of the Royal Astronomical Society 501(2), 2983.
- Freudenburg, J. K. C., Givans, J. J., Choi, A., Hirata, C. M., Bennett, C., Cheung, S., Cillis, A., Cottingham, D., Hill, R. J., Mah, J. et al. 2020,
Publications of the Astronomical Society of the Pacific 132(1013), 74504.
- Friedrich, O., Andrade-Oliveira, F., Camacho, H., Alves, O., Rosenfeld, R., Sanchez, J., Fang, X., Eifler, T. F., Krause, E., Chang, C. et al. 2021,
Monthly Notices of the Royal Astronomical Society 503(3), 3125.
- Friedmann, A. 1922,
Zeitschrift für Physik 10, 377.
- Friedmann, A. 1924,
Zeitschrift für Physik 21, 326.
- Fu, L., Kilbinger, M., Erben, T., Heymans, C., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Miller, L., Sembolini, E. et al. 2014,
Monthly Notices of the Royal Astronomical Society 441(3), 2725.

- Fuji, K. 2021,
<https://github.com/fujiisoup/py3nj>.
- Furlanetto, S. R., Peng Oh, S. & Briggs, F. H. 2006,
Physics Reports 433(4), 181.
- Gamow, G. 1948,
Physical Review 74(4), 505.
- Gamow, G. 1948,
Nature 162(4122), 680.
- García-García, C., Alonso, D. & Bellini, E. 2019,
Journal of Cosmology and Astroparticle Physics 2019(11), 043.
- Gaztanaga, E., Schmidt, S. J., Schneider, M. D. & Tyson, J. A. 2021,
Monthly Notices of the Royal Astronomical Society 503(4), 4964.
- George, D. & Huerta, E. A. 2018,
Physical Review D 97(4), 44039.
- Giblin, B., Cataneo, M., Moews, B. & Heymans, C. 2019,
Monthly Notices of the Royal Astronomical Society 490(4), 4826.
- Giblin, B., Heymans, C., Asgari, M., Hildebrandt, H., Hoekstra, H., Joachimi, B., Kan-nawadi, A., Kuijken, K., Lin, C.-A., Miller, L. et al. 2021,
Astronomy & Astrophysics 645.
- Gilbertson, W., Nomerotski, A. & Takacs, P. 2017,
Journal of Instrumentation 12(09), C09009.
- Gilman, D., Birrer, S., Nierenberg, A., Treu, T., Du, X. & Benson, A. 2020,
Monthly Notices of the Royal Astronomical Society 491(4), 6077.
- Gladders, M. D. & Yee, H. K. C. 2001,
Astronomical Society of the Pacific Conference Series 232, 126.
- Glorot, X. & Bengio, Y. 2010,
Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics 9, 249.
- Good, I. J. 1963,
Journal of the Royal Statistical Society. Series B (Methodological) 25(2), 377.
- Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M. & Bartelmann, M. 2005,
The Astrophysical Journal 622(2), 759.
- Gouyou Beauchamps, S., Lacasa, F., Tutusaus, I., Aubert, M., Baratta, P., Gorce, A. & Sakr, Z. 2022,
Astronomy & Astrophysics, in press (arXiv:2109.02308).
- Graham, M. L., Connolly, A. J., Ivezić, Ž., Schmidt, S. J., Jones, R. L., Jurić, M., Daniel, S. F. & Yoachim, P. 2018,
The Astronomical Journal 155(1), 1.
- Griffiths, R. 1990,

- Hubble Space Telescope: Wide field and planetary camera instrument handbook*, Space Telescope Science Institute.
- Guth, A. H. & Pi, S.-Y. 1982,
Physical Review E 49(15), 1110.
- Guth, A. H. 1981,
Physical Review D 23(2), 347.
- Guy, L. P., Cuillandre, J.-C., Bachelet, E., Banerji, M., Bauer, F. E., Collett, T., Conselice, C. J., Eggl, S., Ferguson, A., Fontana, A. et al. 2022,
 preprint (arXiv:2201.03862).
- Hall, A., Bonvin, C. & Challinor, A. 2013,
Physical Review D 87(6), 64026.
- Hall, A. & Taylor, A. 2019,
Monthly Notices of the Royal Astronomical Society 483(1), 189.
- Hall, A. & Taylor, A. 2022,
Physical Review D 105(12), 123527.
- Hamana, T., Miyazaki, S., Shimasaku, K., Furusawa, H., Doi, M., Hamabe, M., Imi, K., Kimura, M., Komiyama, Y., Nakata, F. et al. 2003,
The Astrophysical Journal 597(1), 98.
- Hamilton, A. J., Rimes, C. D. & Scoccimarro, R. 2006,
Monthly Notices of the Royal Astronomical Society 371(3), 1188.
- Hamimeche, S. & Lewis, A. 2008,
Physical Review D 77(10), 103013.
- Hamimeche, S. & Lewis, A. 2009,
Physical Review D 79(8), 083012.
- Hansen, F. K. & Górski, K. M. 2003,
Monthly Notices of the Royal Astronomical Society 343(2), 559.
- Hartlap, J., Schrabback, T., Simon, P. & Schneider, P. 2009,
Astronomy & Astrophysics 504(3), 689.
- Harnois-Déraps, J. & Van Waerbeke, L. 2015,
Monthly Notices of the Royal Astronomical Society 450(3), 2857.
- Harrison, I., Camera, S., Zuntz, J. & Brown, M. L. 2016,
Monthly Notices of the Royal Astronomical Society 463(4), 3674.
- Harnois-Déraps, J., Amon, A., Choi, A., Demchenko, V., Heymans, C., Kannawadi, A., Nakajima, R., Sirks, E., van Waerbeke, L., Cai, Y. C. et al. 2018,
Monthly Notices of the Royal Astronomical Society 481(1), 1337.
- Harnois-Déraps, J., Giblin, B. & Joachimi, B. 2019,
Astronomy & Astrophysics 631, A160.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J. et al. 2020,
Nature 585(7825), 357.

- Harrison, I., Brown, M. L., Tunbridge, B., Thomas, D. B., Hillier, T., Thomson, A. P., Whitaker, L., Abdalla, F. B., Battye, R. A., Bonaldi, A. et al. 2020,
Monthly Notices of the Royal Astronomical Society 495(2), 1737.
- Harnois-Déraps, J., Martinet, N. & Reischke, R. 2022,
Monthly Notices of the Royal Astronomical Society 509(3), 3868.
- Hazumi, M., Ade, P. A. R., Akiba, Y., Alonso, D., Arnold, K., Aumont, J., Baccigalupi, C., Barron, D., Basak, S., Beckman, S. et al. 2019,
Journal of Low Temperature Physics 194(5-6), 443.
- Hearin, A. P., Zentner, A. R., Ma, Z. & Huterer, D. 2010,
The Astrophysical Journal 720(2), 1351.
- Heavens, A. 2003,
Monthly Notices of the Royal Astronomical Society 343(4), 1327.
- Henghes, B., Pettitt, C., Thiyagalingam, J., Hey, T. & Lahav, O. 2021,
Monthly Notices of the Royal Astronomical Society 505(4), 4847.
- Heymans, C. & Heavens, A. 2003,
Monthly Notices of the Royal Astronomical Society 339(3), 711.
- Heymans, C., Brown, M., Heavens, A., Meisenheimer, K., Taylor, A. & Wolf, C. 2004,
Monthly Notices of the Royal Astronomical Society 347(3), 895.
- Heymans, C., Brown, M. L., Barden, M., Caldwell, J. A. R., Jahnke, K., Peng, C. Y., Rix, H.-W., Taylor, A., Beckwith, S. V. W., Bell, E. F. et al. 2005,
Monthly Notices of the Royal Astronomical Society 361(1), 160.
- Heymans, C., Van Waerbeke, L., Miller, L., Erben, T., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Simon, P., Bonnett, C. et al. 2012,
Monthly Notices of the Royal Astronomical Society 427(1), 146.
- Heymans, C., Grocott, E., Heavens, A., Kilbinger, M., Kitching, T. D., Simpson, F., Benjamin, J., Erben, T., Hildebrandt, H., Hoekstra, H. et al. 2013,
Monthly Notices of the Royal Astronomical Society 432(3), 2433.
- Heymans, C., Tröster, T., Asgari, M., Blake, C., Hildebrandt, H., Joachimi, B., Kuijken, K., Lin, C.-A., Sánchez, A. G., van den Busch, J. L. et al. 2021,
Astronomy & Astrophysics 646, A140.
- Hezaveh, Y., Dalal, N., Holder, G., Kisner, T., Kuhlen, M. & Levasseur, L. P. 2016,
Journal of Cosmology and Astroparticle Physics 2016(11), 48.
- Hezaveh, Y. D., Levasseur, L. P. & Marshall, P. J. 2017,
Nature 548(7669), 555.
- Hikage, C., Oguri, M., Hamana, T., More, S., Mandelbaum, R., Takada, M., Köhlinger, F., Miyatake, H., Nishizawa, A. J., Aihara, H. et al. 2019,
Publications of the Astronomical Society of Japan 71(2), 43.
- Hilbert, S., Hartlap, J. & Schneider, P. 2011,
Astronomy & Astrophysics 536, A85.
- Hildebrandt, H., Viola, M., Heymans, C., Joudaki, S., Kuijken, K., Blake, C., Erben, T., Joachimi, B., Klaes, D., Miller, L. et al. 2017,

- Monthly Notices of the Royal Astronomical Society 465(2), 1454.
 Hillier, T., Brown, M. L., Harrison, I. & Whittaker, L. 2019,
 Monthly Notices of the Royal Astronomical Society 488(4), 5420.
 Hillier, T. J. M. 2020,
 PhD thesis, University of Manchester.
 Hirata, C. M. & Seljak, U. 2004,
 Physical Review D 70(6), 063526.
 Hivon, E., Górski, K. M., Barth Netterfield, C., Crill, B. P., Prunet, S. & Hansen, F. 2002,
 The Astrophysical Journal 567(1), 2.
 Hoekstra, H., Yee, H. K. C. & Gladders, M. D. 2002,
 The Astrophysical Journal 577(2), 595.
 Hoekstra, H., Yee, H. K. C., Gladders, M. D., Barrientos, L. F., Hall, P. B. & Infante, L. 2002,
 The Astrophysical Journal 572(1), 55.
 Hoekstra, H., Kannawadi, A. & Kitching, T. D. 2021,
 Astronomy & Astrophysics 646, A124.
 Howlett, C., Lewis, A., Hall, A. & Challinor, A. 2012,
 Journal of Cosmology and Astroparticle Physics 2012(04), 027.
 Howlett, C., Ross, A. J., Samushia, L., Percival, W. J. & Manera, M. 2015,
 Monthly Notices of the Royal Astronomical Society 449(1), 848.
 Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M.
 & Adam, H. 2017,
 preprint (arXiv:1704.04861).
 Hoyle, B. 2016,
 Astronomy and Computing 16, 34.
 Huang, Y., Addison, G. & Bennett, C. 2019,
 The Astrophysical Journal 882(2), 124.
 Huang, H.-J., Eifler, T., Mandelbaum, R., Bernstein, G. M., Chen, A., Choi, A., García-
 Bellido, J., Huterer, D., Krause, E., Rozo, E. et al. 2021,
 Monthly Notices of the Royal Astronomical Society 502(4), 6010.
 Hubble, E. 1929,
 Proceedings of the National Academy of Science 15(3), 168.
 Hui, H., Ade, P. A. R., Ahmed, Z., Aikin, R., Alexander, K. D., Barkats, D., Benton, S. J.,
 Bischoff, C. A., Bock, J. J., Bowens-Rubin, R. et al. 2018,
 Proceedings of the Society of Photo-Optical Instrumentation Engineers 10708, 49.
 Hunter, J. D. 2007,
 Computing in Science & Engineering 9(3), 90.
 Huterer, D. & Takada, M. 2005,
 Astroparticle Physics 23(4), 369.
 Ivezić, Z., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad,
 Y., Anderson, S. F., Andrew, J. et al. 2019,
 The Astrophysical Journal 873(2), 111.

- Jacobs, C., Glazebrook, K., Collett, T., More, A. & McCarthy, C. 2017,
Monthly Notices of the Royal Astronomical Society 471(1), 167.
- Jarvis, M., Bernstein, G. M., Fischer, P., Smith, D., Jain, B., Tyson, J. A. & Wittman, D. 2003,
The Astronomical Journal 125(3), 1014.
- Jarvis, M., Bernstein, G. & Jain, B. 2004,
Monthly Notices of the Royal Astronomical Society 352(1), 338.
- Jarvis, M. & Jain, B. 2004,
preprint (arXiv:astro-ph/0412234).
- Jing, Y. P., Zhang, P., Lin, W. P., Gao, L. & Springel, V. 2006,
The Astrophysical Journal 640(2), L119.
- Joachimi, B. & Schneider, P. 2008,
Astronomy & Astrophysics 488(3), 829.
- Joachimi, B., Schneider, P. & Eifler, T. 2008,
Astronomy & Astrophysics 477(1), 43.
- Joachimi, B., Schneider, P. & Eifler, T. 2008,
Astronomy & Astrophysics 477(1), 43.
- Joachimi, B. & Bridle, S. L. 2010,
Astronomy & Astrophysics 523, A1.
- Joachimi, B., Mandelbaum, R., Abdalla, F. B. & Bridle, S. L. 2011,
Astronomy & Astrophysics 527, A26.
- Joachimi, B., Lin, C.-A., Asgari, M., Tröster, T., Heymans, C., Hildebrandt, H., Köhlinger, F.,
Sánchez, A. G., Wright, A. H., Bilicki, M. et al. 2021,
Astronomy & Astrophysics 646, A129.
- Jones, R. et al. 2021,
<https://github.com/r1chardj0n3s/parse>.
- Joudaki, S., Blake, C., Heymans, C., Choi, A., Harnois-Deraps, J., Hildebrandt, H., Joachimi,
B., Johnson, A., Mead, A., Parkinson, D. et al. 2017,
Monthly Notices of the Royal Astronomical Society 465(2), 2033.
- Kaiser, N., Wilson, G. & Luppino, G. A. 2000,
preprint (arXiv:astro-ph/0003338).
- Kalus, B., Percival, W. J. & Samushia, L. 2016,
Monthly Notices of the Royal Astronomical Society 455(3), 2573.
- Kamionkowski, M. & Kovetz, E. D. 2016,
Annual Review of Astronomy and Astrophysics 54(1), 227.
- Kamionkowski, M. & Loeb, A. 1997,
Physical Review D 56(8), 4511.
- Kapteyn, J. C. 1922,
The Astrophysical Journal 55, 302.
- Kawamura, S., Ando, M., Seto, N., Sato, S., Musha, M., Kawano, I., Yokoyama, J., Tanaka,
T., Ioka, K., Akutsu, T. et al. 2021,
Progress of Theoretical and Experimental Physics 2021(5), 05A105.

- Kazin, E. A., Koda, J., Blake, C., Padmanabhan, N., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S., Croton, D. J. et al. 2014,
Monthly Notices of the Royal Astronomical Society 441(4), 3524.
- Keitel, D. & Schneider, P. 2011,
Astronomy & Astrophysics 534, A76.
- Kermish, Z. D., Ade, P., Anthony, A., Arnold, K., Barron, D., Boettger, D., Borrill, J., Chapman, S., Chinone, Y., Dobbs, M. A. et al. 2012,
Proceedings of the Society of Photo-Optical Instrumentation Engineers 8452, 366.
- Khetan, N., Izzo, L., Branchesi, M., Wojtak, R., Cantiello, M., Murugeshan, C., Agnello, A., Cappellaro, E., Della Valle, M., Gall, C. et al. 2021,
Astronomy & Astrophysics 647, A72.
- Kilbinger, M., Fu, L., Heymans, C., Simpson, F., Benjamin, J., Erben, T., Harnois-Déraps, J., Hoekstra, H., Hildebrandt, H., Kitching, T. D. et al. 2013,
Monthly Notices of the Royal Astronomical Society 430(3), 2200.
- Kilbinger, M., Heymans, C., Asgari, M., Joudaki, S., Schneider, P., Simon, P., Van Waerbeke, L., Harnois-Déraps, J., Hildebrandt, H., Köhlinger, F. et al. 2017,
Monthly Notices of the Royal Astronomical Society 472(2), 2126.
- Kim, E. J. & Brunner, R. J. 2017,
Monthly Notices of the Royal Astronomical Society 464(4), 4463.
- Kingma, D. P. & Ba, J. 2014,
preprint (arXiv:1412.6980).
- Kirshner, R. P. 2004,
Proceedings of the National Academy of Sciences 101(1), 8.
- Kitching, T. D. & Taylor, A. N. 2011,
Monthly Notices of the Royal Astronomical Society 416(3), 1717.
- Kitching, T. D., Heavens, A. F., Alsing, J., Erben, T., Heymans, C., Hildebrandt, H., Hoekstra, H., Jaffe, A., Kiessling, A., Mellier, Y. et al. 2014,
Monthly Notices of the Royal Astronomical Society 442(2), 1326.
- Kodwani, D., Alonso, D. & Ferreira, P. 2019,
The Open Journal of Astrophysics 2, 1811.11584.
- Kogut, A., Spergel, D. N., Barnes, C., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Limon, M., Meyer, S. S., Page, L. et al. 2003,
The Astrophysical Journal Supplement Series 148(1), 161.
- Kraskov, A., Stögbauer, H. & Grassberger, P. 2004,
Physical Review E 69(6), 066138.
- Krause, E., Eifler, T. F., Zuntz, J., Friedrich, O., Troxel, M. A., Dodelson, S., Blazek, J., Secco, L. F., MacCrann, N., Baxter, E. et al. 2017,
preprint (arXiv:1706.09359).
- Krause, E. & Eifler, T. 2017,
Monthly Notices of the Royal Astronomical Society 470(2), 2100.

- Kramer, M., Stairs, I. H., Manchester, R. N., Wex, N., Deller, A. T., Coles, W. A., Ali, M., Burgay, M., Camilo, F., Cognard, I. et al. 2021,
Physical Review X 11(4), 41050.
- Krause, E., Fang, X., Pandey, S., Secco, L., Alves, O., Huang, H., Blazek, J., Prat, J., Zuntz, J., Eifler, T. et al. 2021,
preprint (arXiv:2105.13548).
- Kuijken, K., Heymans, C., Dvornik, A., Hildebrandt, H., de Jong, J. T. A., Wright, A. H., Erben, T., Bilicki, M., Giblin, B., Shan, H.-Y. et al. 2019,
Astronomy & Astrophysics 625.
- Kuijken, K. 2011,
The Messenger 146, 8.
- Kuijken, K. 1999,
Astronomy & Astrophysics 352, 355.
- Laureijs, R., Amiaux, J., Arduini, S., Auguères, J. .--L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A. et al. 2011,
preprint (arXiv:1110.3193).
- Lee, J., Xiao, L., Schoenholz, S. S., Bahri, Y., Novak, R., Sohl-Dickstein, J. & Pennington, J. 2020,
Journal of Statistical Mechanics: Theory and Experiment 2020(12), 124002.
- Leistedt, B., Hogg, D. W., Wechsler, R. H. & DeRose, J. 2019,
The Astrophysical Journal 881(1), 80.
- Lemaître, G. 1927,
Annales de la Société Scientifique de Bruxelles 47, 49.
- Lewis, A., Challinor, A. & Lasenby, A. 2000,
The Astrophysical Journal 538(2), 473.
- Lewis, A., Challinor, A. & Turok, N. 2001,
Physical Review D 65(2), 023505.
- Lewis, A. & Challinor, A. 2006,
Physics Reports 429(1), 1.
- Li, Y., Hu, W. & Takada, M. 2014,
Physical Review D 89(8), 83519.
- Li, R., Frenk, C. S., Cole, S., Gao, L., Bose, S. & Hellwing, W. A. 2016,
Monthly Notices of the Royal Astronomical Society 460(1), 363.
- Lima, E. V. R., Sodré, L., Bom, C. R., Teixeira, G. S. M., Nakazono, L., Buzzo, M. L., Queiroz, C., Herpich, F. R., Castellon, J. L. N., Dantas, M. L. L. et al. 2022,
Astronomy and Computing 38, 100510.
- Lin, C.-H., Harnois-Déraps, J., Eifler, T., Pospisil, T., Mandelbaum, R., Lee, A. B. & Singh, S. 2020,
Monthly Notices of the Royal Astronomical Society 499(2), 2977.
- Linde, A. D. 1982,
Physics Letters B 108(6), 389.

- Linde, A. D. 1983,
 Physics Letters B 129(3), 177.
- Liu, J. & Madhavacheril, M. S. 2019,
 Physical Review D 99(8), 083508.
- Long, X. 2020,
 https://keras.io/examples/vision/super_resolution_sub_pixel.
- Louca, A. J. & Sellentin, E. 2020,
 The Open Journal of Astrophysics 3, 2007.07253.
- Loureiro, A., Whittaker, L., Spurio Mancini, A., Joachimi, B., Cuceu, A., Asgari, M., Stölzner, B., Tröster, T., Wright, A. H., Bilicki, M. et al. 2021,
 preprint (arXiv:2110.06947).
- Lu, T., Zhang, J., Dong, F., Li, Y., Liu, D., Fu, L., Li, G. & Fan, Z. 2017,
 The Astronomical Journal 153(4), 197.
- Ma, S. & Yunes, N. 2019,
 Physical Review D 100(12), 124032.
- Macaulay, E., Wehus, I. K. & Eriksen, H. K. 2013,
 Physical Review Letters 111(16), 161301.
- MacCrann, N., Aleksić, J., Amara, A., Bridle, S. L., Bruderer, C., Chang, C., Dodelson, S., Eifler, T. F., Huff, E. M., Huterer, D. et al. 2017,
 Monthly Notices of the Royal Astronomical Society 465(3), 2567.
- Macaulay, E., Nichol, R. C., Bacon, D., Brout, D., Davis, T. M., Zhang, B., Bassett, B. A., Scolnic, D., Möller, A., D'Andrea, C. B. et al. 2019,
 Monthly Notices of the Royal Astronomical Society 486(2), 2184.
- Maggiore, M., Broeck, C. V. D., Bartolo, N., Belgacem, E., Bertacca, D., Bizouard, M. A., Branchesi, M., Clesse, S., Foffa, S., Garc\'ia-Bellido, J. et al. 2020,
 Journal of Cosmology and Astroparticle Physics 2020(03), 50.
- Mahalanobis, P. C. 1936,
 Proceedings of the National Institute of Sciences of India 2(1), 49.
- Mandelbaum, R., Blake, C., Bridle, S., Abdalla, F. B., Brough, S., Colless, M., Couch, W., Croom, S., Davis, T., Drinkwater, M. J. et al. 2011,
 Monthly Notices of the Royal Astronomical Society 410(2), 844.
- Mangilli, A., Plaszczynski, S. & Tristram, M. 2015,
 Monthly Notices of the Royal Astronomical Society 453(3), 3175.
- Manrique-Yus, A. & Sellentin, E. 2020,
 Monthly Notices of the Royal Astronomical Society 491(2), 2655.
- Markevitch, M., Gonzalez, A. H., Clowe, D., Vikhlinin, A., Forman, W., Jones, C., Murray, S. & Tucker, W. 2004,
 The Astrophysical Journal 606(2), 819.
- Martinelli, M., Tatusaus, I., Archidiacono, M., Camera, S., Cardone, V. F., Clesse, S., Casas, S., Casarini, L., Mota, D. F., Hoekstra, H. et al. 2021,
 Astronomy & Astrophysics 649, A100.

- Martel, H., Shapiro, P. R. & Weinberg, S. 1998,
 The Astrophysical Journal 492(1), 29.
- McKinney, W. 2010,
 Proceedings of the 9th Python in Science Conference, 56.
- Melchior, P., Joseph, R., Sanchez, J., MacCrann, N. & Gruen, D. 2021,
 Nature Reviews Physics 3(10), 712.
- Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., Kumar, A.,
 Ivanov, S., Moore, J. K., Singh, S. et al. 2017,
 PeerJ Computer Science 3, e103.
- Miyazaki, S., Komiyama, Y., Sekiguchi, M., Okamura, S., Doi, M., Furusawa, H., Hamabe,
 M., Imi, K., Kimura, M., Nakata, F. et al. 2002,
 Publications of the Astronomical Society of Japan 54(6), 833.
- Miyazaki, S., Komiyama, Y., Nakaya, H., Kamata, Y., Doi, Y., Hamana, T., Karoji, H., Furuya,
 H., Kawanomoto, S., Morokuma, T. et al. 2012,
 Proceedings of the Society of Photo-Optical Instrumentation Engineers 8446, 84460Z.
- Mohammed, I., Martizzi, D., Teyssier, R. & Amara, A. 2014,
 preprint (arXiv:1410.6826).
- Mohan, D., Scaife, A. M. M., Porter, F., Walmsley, M. & Bowles, M. 2022,
 Monthly Notices of the Royal Astronomical Society 511(3), 3722.
- Möller, A. & de Boissière, T. 2020,
 Monthly Notices of the Royal Astronomical Society 491(3), 4277.
- More, S., Miyatake, H., Mandelbaum, R., Takada, M., Spergel, D. N., Brownstein, J. R. &
 Schneider, D. P. 2015,
 The Astrophysical Journal 806(1), 2.
- Moster, B. P., Somerville, R. S., Newman, J. A. & Rix, H.-W. 2011,
 The Astrophysical Journal 731(2), 113.
- Nibali, A., He, Z., Morgan, S. & Prendergast, L. 2018,
 preprint (arXiv:1801.07372).
- Nicolis, A., Rattazzi, R. & Trincherini, E. 2009,
 Physical Review D 79(6), 64036.
- Nicola, A., García-García, C., Alonso, D., Dunkley, J., Ferreira, P. G., Slosar, A. & Spergel,
 D. N. 2021,
 Journal of Cosmology and Astroparticle Physics 2021(03), 067.
- Niu, Z. M. & Liang, H. Z. 2018,
 Physics Letters B 778, 48.
- Nojiri, S. & Odintsov, S. D. 2003,
 Physical Review D 68(12), 123512.
- Nojiri, S. & Odintsov, S. D. 2011,
 Physics Reports 505(2), 59.
- Nojiri, S., Odintsov, S. D. & Oikonomou, V. K. 2017,
 Physics Reports 692, 1.

- Nourbakhsh, E., Tyson, J. A., Schmidt, S. J. & The LSST Dark Energy Science Collaboration 2022,
Monthly Notices of the Royal Astronomical Society 514(4), 5905.
- Nurbaeva, G., Tewes, M., Courbin, F. & Meylan, G. 2015,
Astronomy & Astrophysics 577, A104.
- O’Raifeartaigh, C. & Mitton, S. 2018,
Physics in Perspective 20, 318.
- Pasquet, J., Bertin, E., Treyer, M., Arnouts, S. & Fouchez, D. 2019,
Astronomy & Astrophysics 621, A26.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. 2011,
Journal of Machine Learning Research 12, 2825.
- Peebles, P. J. E. 1982,
The Astrophysical Journal 263, L1.
- Penzias, A. A. & Wilson, R. W. 1965,
The Astrophysical Journal 142, 419.
- Pence, W. 1999,
Astronomical Data Analysis Software and Systems VIII 172, 487.
- Percival, W. J. & Brown, M. L. 2006,
Monthly Notices of the Royal Astronomical Society 372(3), 1104.
- Percival, W. J. & White, M. 2009,
Monthly Notices of the Royal Astronomical Society 393(1), 297.
- Percival, W. J., Reid, B. A., Eisenstein, D. J., Bahcall, N. A., Budavari, T., Frieman, J. A., Fukugita, M., Gunn, J. E., Ivezić, Ž., Knapp, G. R. et al. 2010,
Monthly Notices of the Royal Astronomical Society 401(4), 2148.
- Percival, W. J., Samushia, L., Ross, A. J., Shapiro, C. & Raccanelli, A. 2011,
Philosophical Transactions of the Royal Society of London Series A 369(1957), 5058.
- Perreault Levasseur, L., Hezaveh, Y. D. & Wechsler, R. H. 2017,
The Astrophysical Journal 850(1), L7.
- Perraudin, N., Defferrard, M., Kacprzak, T. & Sgier, R. 2019,
Astronomy and Computing 27, 130.
- Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E. et al. 1999,
The Astrophysical Journal 517(2), 565.
- Petri, A., Haiman, Z. & May, M. 2016,
Physical Review D 93(6), 063524.
- Petrillo, C. E., Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., Napolitano, N. R., Covone, G., Schneider, P., Grado, A. et al. 2017,
Monthly Notices of the Royal Astronomical Society 472(1), 1129.
- Petri, A. 2016,
Astronomy and Computing 17, 73.

- Pielorz, J., Rödiger, J., Tereno, I. & Schneider, P. 2010,
Astronomy & Astrophysics 514, A79.
- Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C.,
 Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N. et al. 2019,
Astronomy & Astrophysics 641, A5.
- Planck Collaboration, Aghanim, N., Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Bac-
 cagalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B. et al. 2020,
Astronomy & Astrophysics 641, A1.
- Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C.,
 Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N. et al. 2020,
Astronomy & Astrophysics 641, A6.
- Planck Collaboration, Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C.,
 Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N. et al. 2020,
Astronomy & Astrophysics 641, A10.
- Pober, J. C., Liu, A., Dillon, J. S., Aguirre, J. E., Bowman, J. D., Bradley, R. F., Carilli, C. L.,
 DeBoer, D. R., Hewitt, J. N., Jacobs, D. C. et al. 2014,
The Astrophysical Journal 782(2), 66.
- Pritchard, J. R. & Loeb, A. 2012,
Reports on Progress in Physics 75(8), 86901.
- Pyo, J., Duan, H., Baek, S., Kim, M. S., Jeon, T., Kwon, Y. S., Lee, H. & Cho, K. H. 2019,
Remote Sensing of Environment 233, 111350.
- Raccanelli, A., Bertacca, D., Pietrobon, D., Schmidt, F., Samushia, L., Bartolo, N., Doré, O.,
 Matarrese, S. & Percival, W. J. 2013,
Monthly Notices of the Royal Astronomical Society 436(1), 89.
- Rau, M. M., Morrison, C. B., Schmidt, S. J., Wilson, S., Mandelbaum, R. & Mao, Y.-Y. 2022,
Monthly Notices of the Royal Astronomical Society 509(4), 4886.
- Reddi, S. J., Kale, S. & Kumar, S. 2018,
Proceedings of the International Conference on Learning Representations.
- Refregier, A. & The DUNE Collaboration 2009,
Experimental Astronomy 23(1), 17.
- Reid, M. J., Pesce, D. W. & Riess, A. G. 2019,
The Astrophysical Journal 886(2), L27.
- Rhodes, J., Refregier, A. & Groth, E. J. 2001,
The Astrophysical Journal 552(2), L85.
- Rhodes, J. D., Massey, R. J., Albert, J., Collins, N., Ellis, R. S., Heymans, C., Gardner, J. P.,
 Kneib, J.-P., Koekemoer, A., Leauthaud, A. et al. 2007,
The Astrophysical Journal Supplement Series 172(1), 203.
- Rhodes, J., Leauthaud, A., Stoughton, C., Massey, R., Dawson, K., Kolbe, W. & Roe, N. 2010,
Publications of the Astronomical Society of the Pacific 122(890), 439.
- Rhodes, J., Nichol, R. C., Aubourg, É., Bean, R., Boutigny, D., Bremer, M. N., Capak, P.,
 Cardone, V., Carry, B., Conselice, C. J. et al. 2017,

- The Astrophysical Journal Supplement Series 233(2), 21.
- Ribli, D., Dobos, L. & Csabai, I. 2019,
 Monthly Notices of the Royal Astronomical Society 489(4), 4847.
- Ribli, D., Pataki, B. Á., Zorrilla Matilla, J. M., Hsu, D., Haiman, Z. & Csabai, I. 2019,
 Monthly Notices of the Royal Astronomical Society 490(2), 1843.
- Ribeiro, W. J. M. 2019,
 preprint (arXiv:1904.11068).
- Riess, A. G., Casertano, S., Yuan, W., Macri, L., Anderson, J., MacKenty, J. W., Bowers, J. B.,
 Clubb, K. I., Filippenko, A. V., Jones, D. O. et al. 2018,
 The Astrophysical Journal 855(2), 136.
- Riess, A. G., Casertano, S., Yuan, W., Macri, L., Bucciarelli, B., Lattanzi, M. G., MacKenty,
 J. W., Bowers, J. B., Zheng, W., Filippenko, A. V. et al. 2018,
 The Astrophysical Journal 861(2), 126.
- Riess, A. G., Rodney, S. A., Scolnic, D. M., Shafer, D. L., Strolger, L.-G., Ferguson, H. C.,
 Postman, M., Graur, O., Maoz, D., Jha, S. W. et al. 2018,
 The Astrophysical Journal 853(2), 126.
- Riess, A. G., Yuan, W., Casertano, S., Macri, L. M. & Scolnic, D. 2020,
 The Astrophysical Journal 896(2), L43.
- Riess, A. G., Casertano, S., Yuan, W., Bowers, J. B., Macri, L., Zinn, J. C. & Scolnic, D. 2021,
 The Astrophysical Journal Letters 908(1), L6.
- Riess, A. G., Yuan, W., Macri, L. M., Scolnic, D., Brout, D., Casertano, S., Jones, D. O.,
 Murakami, Y., Breuval, L., Brink, T. G. et al. 2022,
 The Astrophysical Journal Letters 934(1), L7.
- Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M.,
 Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P. et al. 1998,
 The Astronomical Journal 116(3), 1009.
- Rivi, M., Miller, L., Makhathini, S. & Abdalla, F. B. 2016,
 Monthly Notices of the Royal Astronomical Society 463(2), 1881.
- Rivi, M. & Miller, L. 2018,
 Monthly Notices of the Royal Astronomical Society 476(2), 2053.
- Rivi, M., Lochner, M., Balan, S. T., Harrison, I. & Abdalla, F. B. 2019,
 Monthly Notices of the Royal Astronomical Society 482(1), 1096.
- Romanelli, M., Riess, A., Mancino, S., Anderson, R. I., Freudling, W., Kudritzki, R.-P.,
 Macri, L., Mucciarelli, A. & Yuan, W. 2022,
 Astronomy & Astrophysics 658, A29.
- Rowlands, N., Midwinter, C. & Warner, G. 2018,
 Proceedings of the Society of Photo-Optical Instrumentation Engineers 10709, 416.
- Rowe, B. 2010,
 Monthly Notices of the Royal Astronomical Society 404(1), 350.
- Safran, I. & Shamir, O. 2016,
 preprint (arXiv:1610.09887).

- Samuroff, S., Mandelbaum, R. & Blazek, J. 2021,
Monthly Notices of the Royal Astronomical Society 508(1), 637.
- Sanchez, J., Mendoza, I., Kirkby, D. P. & Burchat, P. R. 2021,
Journal of Cosmology and Astroparticle Physics 2021(07), 43.
- Sato, M., Ichiki, K. & Takeuchi, T. T. 2010,
Physical Review Letters 105(25), 251301.
- Sato, M., Takada, M., Hamana, T. & Matsubara, T. 2011,
The Astrophysical Journal 734(2), 76.
- Sato, M. & Nishimichi, T. 2013,
Physical Review D 87(12), 123538.
- Saxe, A. M., McClelland, J. L. & Ganguli, S. 2013,
preprint (arXiv:1312.6120).
- Scaife, A. M. M. & Porter, F. 2021,
Monthly Notices of the Royal Astronomical Society 503(2), 2369.
- Schneider, P., van Waerbeke, L., Kilbinger, M. & Mellier, Y. 2002,
Astronomy & Astrophysics 396(1), 1.
- Schneider, M. D. & Bridle, S. 2010,
Monthly Notices of the Royal Astronomical Society 402(4), 2127.
- Schneider, A., Teyssier, R., Potter, D., Stadel, J., Onions, J., Reed, D. S., Smith, R. E., Springel, V., Pearce, F. R. & Scoccimarro, R. 2016,
Journal of Cosmology and Astroparticle Physics 2016(04), 47.
- Schmidt, A. S., White, S. D., Schmidt, F. & Stücker, J. 2018,
Monthly Notices of the Royal Astronomical Society 479(1), 162.
- Schmidt, S. J., Malz, A. I., Soo, J. Y. H., Almosallam, I. A., Brescia, M., Cavaoti, S., Cohen-Tanugi, J., Connolly, A. J., DeRose, J., Freeman, P. E. et al. 2020,
Monthly Notices of the Royal Astronomical Society 499(2), 1587.
- Schneider, A., Stoira, N., Refregier, A., Weiss, A. J., Knabenhans, M., Stadel, J. & Teyssier, R. 2020,
Journal of Cosmology and Astroparticle Physics 2020(04), 19.
- Schuldt, S., Suyu, S. H., Cañameras, R., Taubenberger, S., Meinhardt, T., Leal-Taixé, L. & Hsieh, B. C. 2021,
Astronomy & Astrophysics 651, A55.
- Scoccimarro, R., Zaldarriaga, M. & Hui, L. 1999,
The Astrophysical Journal 527(1), 1.
- Sellentin, E. & Heavens, A. F. 2016,
Monthly Notices of the Royal Astronomical Society 464(4), 4658.
- Sellentin, E. & Heavens, A. F. 2016,
Monthly Notices of the Royal Astronomical Society 456(1), L132.
- Sellentin, E. & Heavens, A. F. 2018,
Monthly Notices of the Royal Astronomical Society 473(2), 2355.
- Sellentin, E., Heymans, C. & Harnois-Déraps, J. 2018,

- Monthly Notices of the Royal Astronomical Society 477(4), 4879.
- Sellentin, E. 2015,
 Monthly Notices of the Royal Astronomical Society 453(1), 893.
- Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P. & McCarthy, I. G. 2011,
 Monthly Notices of the Royal Astronomical Society 417(3), 2020.
- Semboloni, E., Hoekstra, H. & Schaye, J. 2013,
 Monthly Notices of the Royal Astronomical Society 434(1), 148.
- Sgier, R. J., Réfrégier, A., Amara, A. & Nicola, A. 2019,
 Journal of Cosmology and Astroparticle Physics 2019(01), 44.
- Shao, L., Sennett, N., Buonanno, A., Kramer, M. & Wex, N. 2017,
 Physical Review X 7(4), 41025.
- Shirasaki, M., Hamana, T. & Yoshida, N. 2015,
 Monthly Notices of the Royal Astronomical Society 453(3), 3043.
- Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D. & Wang, Z. 2016,
 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1874.
- Singh, S., Mandelbaum, R. & More, S. 2015,
 Monthly Notices of the Royal Astronomical Society 450(2), 2195.
- Sklar, A. 1959,
 Publications de l'Institut Statistique de l'Université de Paris 8, 229.
- Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G. & Couchman, H. M. P. 2003,
 Monthly Notices of the Royal Astronomical Society 341(4), 1311.
- Smith, W. B. & Hocking, R. R. 1972,
 Journal of the Royal Statistical Society. Series C (Applied Statistics) 21(3), 341.
- Soltis, J., Casertano, S. & Riess, A. G. 2021,
 The Astrophysical Journal Letters 908(1), L5.
- Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O. et al. 2015,
 Square Kilometre Array Cosmology Science Working Group, Bacon, D. J., Battye, R. A., Bull, P., Camera, S., Ferreira, P. G., Harrison, I., Parkinson, D., Pourtsidou, A., Santos, M. G. et al. 2020,
 Publications of the Astronomical Society of Australia 37, E007.
- Starobinsky, A. A. 1982,
 Physics Letters B 117(3), 175.
- Stebbins, A. 1996,
 preprint (arXiv:astro-ph/9609149).
- Sun, L., Fan, Z.-H., Tao, C., Kneib, J.-P., Jouvel, S. & Tilquin, A. 2009,
 The Astrophysical Journal 699(2), 958.
- Sun, L., Wang, Q. & Zhan, H. 2013,
 The Astrophysical Journal 777(1), 75.

- Swetz, D. S., Ade, P. A. R., Amiri, M., Appel, J. W., Battistelli, E. S., Burger, B., Chervenak, J., Devlin, M. J., Dicker, S. R., Dorie, W. B. et al. 2011,
The Astrophysical Journal Supplement Series 194(2), 41.
- Takada, M. & Bridle, S. 2007,
New Journal of Physics 9(12), 446.
- Takada, M. & Jain, B. 2009,
Monthly Notices of the Royal Astronomical Society 395(4), 2065.
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A. & Oguri, M. 2012,
The Astrophysical Journal 761(2), 152.
- Takada, M. & Hu, W. 2013,
Physical Review D 87(12), 123504.
- Takahashi, R., Hamana, T., Shirasaki, M., Namikawa, T., Nishimichi, T., Osato, K. & Shiroyama, K. 2017,
The Astrophysical Journal 850(1), 24.
- Taruya, A., Takada, M., Hamana, T., Kayo, I. & Futamase, T. 2002,
The Astrophysical Journal 571(2), 638.
- Taubenberger, S., Suyu, S. H., Komatsu, E., Jee, I., Birrer, S., Bonvin, V., Courbin, F., Rusu, C. E., Shajib, A. J. & Wong, K. C. 2019,
Astronomy & Astrophysics 628, L7.
- Taylor, A. N. & Kitching, T. D. 2010,
Monthly Notices of the Royal Astronomical Society 408(2), 865.
- Taylor, P. L., Bernardeau, F. & Kitching, T. D. 2018,
Physical Review D 98(8), 83514.
- Taylor, P. L., Kitching, T. D., Alsing, J., Wandelt, B. D., Feeney, S. M. & McEwen, J. D. 2019,
Physical Review D 100(2), 023519.
- Taylor, P. L., Bernardeau, F. & Huff, E. 2021,
Physical Review D 103(4), 43531.
- Taylor, P. L., Kitching, T., Cardone, V., Ferté, A., Huff, E., Bernardeau, F., Rhodes, J., Deshpande, A., Tutusaus, I., Pourtsidou, A. et al. 2021,
The Open Journal of Astrophysics 4(1), 6.
- Tegmark, M. 1997,
Physical Review D 55(10), 5895.
- Tessore, N. 2019,
preprint (arXiv:1904.09973).
- Tewes, M., Kuntzer, T., Nakajima, R., Courbin, F., Hildebrandt, H. & Schrabbach, T. 2019,
Astronomy & Astrophysics 621, A36.
- The Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M. et al. 2013,
Astronomy & Astrophysics 558.
- The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N. et al. 2018,

- The Astronomical Journal 156(3), 123.
- The pandas Development Team 2020,
<https://doi.org/10.5281/zenodo.3715232>.
- Trevisani, N. 2018,
Universe 4(11), 131.
- Troxel, M. & Ishak, M. 2015,
Physics Reports 558, 1.
- Troxel, M. A., MacCrann, N., Zuntz, J., Eifler, T. F., Krause, E., Dodelson, S., Gruen, D., Blazek, J., Friedrich, O., Samuroff, S. et al. 2018,
Physical Review D 98(4), 043528.
- Tsujikawa, S. 2013,
Classical and Quantum Gravity 30(21), 214003.
- Utama, R., Piekarewicz, J. & Prosper, H. B. 2016,
Physical Review C 93(1), 14311.
- Vainio, J. & Vilja, I. 2017,
General Relativity and Gravitation 49, 99.
- Van Albada, T., Bahcall, J., Begeman, K. & Sancisi, R. 1985,
The Astrophysical Journal 295, 305.
- Van Waerbeke, L., Mellier, Y., Erben, T., Cuillandre, J. C., Bernardeau, F., Maoli, R., Bertin, E., McCracken, H. J., Le Fèvre, O., Fort, B. et al. 2000,
Astronomy & Astrophysics 358, 30.
- Van Waerbeke, L., Mellier, Y. & Hoekstra, H. 2005,
Astronomy & Astrophysics 429(1), 75.
- Van Waerbeke, L., Benjamin, J., Erben, T., Heymans, C., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Miller, L., Coupon, J. et al. 2013,
Monthly Notices of the Royal Astronomical Society 433(4), 3373.
- Van Daalen, M. P., Schaye, J., Booth, C. M. & Dalla Vecchia, C. 2011,
Monthly Notices of the Royal Astronomical Society 415(4), 3649.
- Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T. & Bolton, A. S. 2014,
Monthly Notices of the Royal Astronomical Society 442(3), 2017.
- Ver Steeg, G. 2014,
https://github.com/gregversteeg/NPEET/blob/master/npeet_doc.pdf.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J. et al. 2020,
Nature Methods 17(3), 261.
- Wagner, C., Schmidt, F., Chiang, C.-T. & Komatsu, E. 2015,
Journal of Cosmology and Astroparticle Physics 2015(8), 42.
- Walter, C. W. 2015,
Journal of Instrumentation 10(05), C05015.
- Wandelt, B. D., Hivon, E. & Górski, K. M. 2001,
Physical Review D 64(8), 083003.

- Wang, M., Percival, W. J., Avila, S., Crittenden, R. & Bianchi, D. 2019,
Monthly Notices of the Royal Astronomical Society 486(1), 951.
- Waskom, M. L. 2021,
Journal of Open Source Software 6(60), 3021.
- Whittaker, L., Brown, M. L. & Battye, R. A. 2015,
Monthly Notices of the Royal Astronomical Society 451(1), 383.
- Wieczorek, M. A. & Meschede, M. 2018,
Geochemistry, Geophysics, Geosystems 19(8), 2574.
- Wilking, P. & Schneider, P. 2013,
Astronomy & Astrophysics 556, A70.
- Wilking, P., Rösler, R. & Schneider, P. 2015,
Astronomy & Astrophysics 582, A107.
- Wiltshire, D. L. 2007,
New Journal of Physics 9(10), 377.
- Wittman, D. M., Tyson, J. A., Kirkman, D., Dell'Antonio, I. & Bernstein, G. 2000,
Nature 405(6783), 143.
- Wittman, D. M., Tyson, J. A., Bernstein, G. M., Lee, R. W., Dell'Antonio, I. P., Fischer, P.,
 Smith, D. R. & Blouke, M. M. 1998,
Proceedings of the Society of Photo-Optical Instrumentation Engineers 3355.
- Wolf, C., Dye, S., Kleinheinrich, M., Meisenheimer, K., Rix, H.-W. & Wisotzki, L. 2001,
Astronomy & Astrophysics 377(2), 442.
- Wright, A. H., Hildebrandt, H., van den Busch, J. L. & Heymans, C. 2020,
Astronomy & Astrophysics 637, A100.
- York, D. G., Adelman, J., John E. Anderson, J., Anderson, S. F., Annis, J., Bahcall, N. A.,
 Bakken, J. A., Barkhouse, R., Bastian, S., Berman, E. et al. 2000,
The Astronomical Journal 120(3), 1579.
- Yuan, W., Riess, A. G., Macri, L. M., Casertano, S. & Scolnic, D. M. 2019,
The Astrophysical Journal 886(1), 61.
- Zacharegkas, G., Chang, C., Prat, J., Pandey, S., Ferrero, I., Blazek, J., Jain, B., Crocce, M.,
 DeRose, J., Palmese, A. et al. 2022,
Monthly Notices of the Royal Astronomical Society 509(3), 3119.
- Zentner, A. R., Semboloni, E., Dodelson, S., Eifler, T., Krause, E. & Hearn, A. P. 2013,
Physical Review D 87(4), 43509.
- Zhang, T., Mandelbaum, R. & The LSST Dark Energy Science Collaboration 2022,
Monthly Notices of the Royal Astronomical Society 510(2), 1978.
- Zonca, A., Singer, L., Lenz, D., Reinecke, M., Rosset, C., Hivon, E. & Gorski, K. 2019,
Journal of Open Source Software 4(35), 1298.
- Zuntz, J., Paterno, M., Jennings, E., Rudd, D., Manzotti, A., Dodelson, S., Bridle, S., Sehrish,
 S. & Kowalkowski, J. 2015,
Astronomy and Computing 12, 45.
- Zwicky, F. 1933,

Helvetica Physica Acta 6, 110.
Zwicky, F. 1937,
The Astrophysical Journal 86, 217.