Higher-order statistics in cosmic shear

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.) der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von **Sven Heydenreich** aus Dortmund

Bonn, August 2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Peter Schneider
2. Gutachter:	Prof. Dr. Cristiano Porciani

Tag der Promotion:	29.11.2022
Erscheinungsjahr:	2023

Abstract

The humble goal of cosmology is to understand the Universe as a whole. In the last decades, the Λ cold dark matter model has proven itself as the standard model of cosmology by explaining the majority of cosmological observations with remarkable accuracy. However, there is growing evidence that this model will need to be revised: Aside from our inability to explain its main components, in recent years, internal tensions between constraints of the model parameters have emerged.

The weak gravitational lensing effect is one of the best tools to constrain the standard model of cosmology and test potential extensions. It describes the deflection of light from distant galaxies by the tidal gravitational fields of the large-scale structure of the Universe and is an excellent means to study the evolution of structure growth, as it allows us to map the dark matter distribution directly.

In this thesis, I study the use of higher-order statistics to improve the constraining power of weak lensing surveys. I test two higher-order statistics: persistent homology and third-order aperture mass statistics. For the former, we develop a simulation-based inference pipeline to constrain cosmological parameters, show that they perform better than peak count statistics, and conduct a cosmological parameter inference on the year-1 data of the Dark Energy Survey, where we find for the matter clustering parameter $S_8 = 0.747^{+0.025}_{-0.031}$, which is in full agreement with our analysis of second-order statistics ($S_8 = 0.759^{+0.049}_{-0.042}$). We further measure a tension in the matter density parameter Ω_m, where persistent homology yields $\Omega_\mathrm{m} = 0.468^{+0.051}_{-0.036}$, in contrast to $\Omega_\mathrm{m} = 0.256^{+0.034}_{-0.058}$ from the shear two-point correlation functions.

In the second part of this thesis, I present our preparations for a cosmological parameter analysis with third-order aperture mass statistics. We develop and test strategies to model them directly from the matter bispectrum, measure them quickly in simulations, and estimate them from real data. The ability to model these statistics directly gives rise to many validation tests, which we perform to ensure that a subsequent cosmological parameter analysis remains unbiased. Using mock data, we show that a combined analysis of second- and third-order shear statistics of a current-generation survey remains unbiased in the absence of systematics and yields an improvement on the S_8-constraints by almost a factor of two, compared to analysis with only second-order statistics. We then develop a

model for the covariance of third-order aperture masses and a method to test the individual emerging terms using N-body simulations.

I conclude this thesis by comparing the simulation-based inference method to the direct modelling approach for higher-order shear statistics. I hope our efforts will contribute to testing the standard model of cosmology and its potential extensions with the next generation of weak lensing surveys.

Contents

CHAPTER 1

Introduction

"It has been said that astronomy is a humbling and character-building experience. There is perhaps no better demonstration of the folly of human conceits than this distant image of our tiny world. To me, it underscores our responsibility to deal more kindly with one another, and to preserve and cherish the pale blue dot, the only home we've ever known."
– Carl Sagan, Pale Blue Dot

Probably since its birth, and certainly for millennia, humankind has wondered about the origins of the world and the cosmos that surrounds it. The stars not only inspired countless myths and legends but also enabled us to predict the seasons and create calendars. Through the lens of astrology, the movements of planets shaped the fate of entire nations. However, until about 400 years ago, the study of the sky was shrouded by mysticism. The arguably most revolutionary astronomical discovery, and the one that enabled the development of modern cosmology, was Newton's law of universal gravitation. Not only is gravitation the primary driver of structure formation in the Universe, but the implications of this discovery were also much grander: The physical laws on Earth are no different from the ones governing the cosmos. The heavens are no outer-worldly magical place; instead, the Earth is an equal (albeit tiny) part of it. Today, we live in a privileged era where the origins of the Cosmos and the emergence of the Earth, life, and humans, are no longer the topic of legends but a quantitative science.

While scientists and philosophers have pondered questions about the origin and nature of the Universe for thousands of years, the field of modern cosmology is a fairly new one. Despite millennia of observations by countless cultures around the Earth, we only learned that the Universe is bigger than our Galaxy about a century ago (Hubble, 1929b). Ever since then, our understanding of the Universe has advanced rapidly. We discovered that the Universe is expanding (Hubble, 1929a) and that it thus must have originated from a tiny, hot and dense region (Lemaître, 1931). Furthermore, we discovered a dark matter component that does not interact with light (Zwicky, 1933; Rubin and Ford, 1970).

Penzias and Wilson (1965) detected the cosmic microwave background (CMB), which confirmed the big bang hypothesis (Dicke et al., 1965). We discovered that while dark matter dominates the structure formation in our Universe (Peebles, 1982), dark energy dominates its total energy budget today (Garnavich et al., 1998). These discoveries led to the Λ cold dark matter model (ΛCDM) model, which is widely accepted today as the standard model of cosmology. This model describes the origin and evolution of the Universe with remarkable accuracy and enables us to explain a broad set of observations, like the primordial abundances of helium (Peebles, 1966) and other elements (Wagoner et al., 1967), the statistical properties of the CMB, the existence and scale of baryon acoustic oscillations (BAOs) and many more.

Taking a step back, one can not help but feel a sense of awe when contemplating our current picture of the Universe. All of humanity lives on a tiny planet orbiting a relatively average star among the hundreds of billions of other stars in our Galaxy. This galaxy, again, is just one of the hundreds of billions of others that make up the observable Universe. Taking a look at the thousands of background galaxies depicted in the recently released image from the James Webb Space Telescope (Fig. 1.1), and considering that thirty million similar images cover the full sky, one is unable to comprehend the vastness of the Universe that surrounds us. And then, everything in that Universe moves through a four-dimensional spacetime manifold and simultaneously deforms that manifold through its mass (or energy). It is truly a wonderful time to be alive.

Nowadays, as we enter the age of precision cosmology, some tensions seem to be emerging between observations of the distant CMB and the more local Universe. The most prominent one is certainly the tension in the Hubble constant H_0, which describes the Universe's current expansion rate. While Planck Collaboration (2020b) report a value of $H_0 = (67.4 \pm 0.5)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, Riess et al. (2022) measured the same parameter using local supernovae and found $H_0 = (73.04 \pm 1.04)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (for a review, see Di Valentino et al., 2021a). A less prominent tension is the one in the matter clustering parameter $S_8 = \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$, where Ω_m is the matter density parameter and σ_8 describes the normalisation of the power spectrum. Whether we measure this parameter with weak gravitational lensing or galaxy clustering, we find that the Universe is a lot less clustered than observations of the CMB suggest (for a review, see Di Valentino et al., 2021b). While we can not yet rule out that these tensions are mere statistical fluctuations or the result of unknown systematic biases (or a combination of both), this seems increasingly unlikely. In the next decade, future experiments like the Vera Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezic et al., 2008) or *Euclid* (Laureijs et al., 2011) have the potential to confirm the S_8 tension and revolutionise our understanding of the Universe (Abdalla et al., 2022).

Arguably, the nature of dark energy poses one of the biggest mysteries in modern astronomy. This strange component causes an accelerated expansion of our Universe. Several groups (Albrecht et al., 2006; Peacock et al., 2006) have studied how we can test various models of dark energy and agreed on four methods showing promise: (1) measuring the scale of baryonic acoustic oscillations over cosmic time, (2) measuring the number density of galaxy clusters as a function of mass and redshift and (3) using supernovae of type Ia as standard candles to study the expansion history of the Universe. The most promising method, (4), is called *cosmic shear* and focuses on measuring the distortion of background

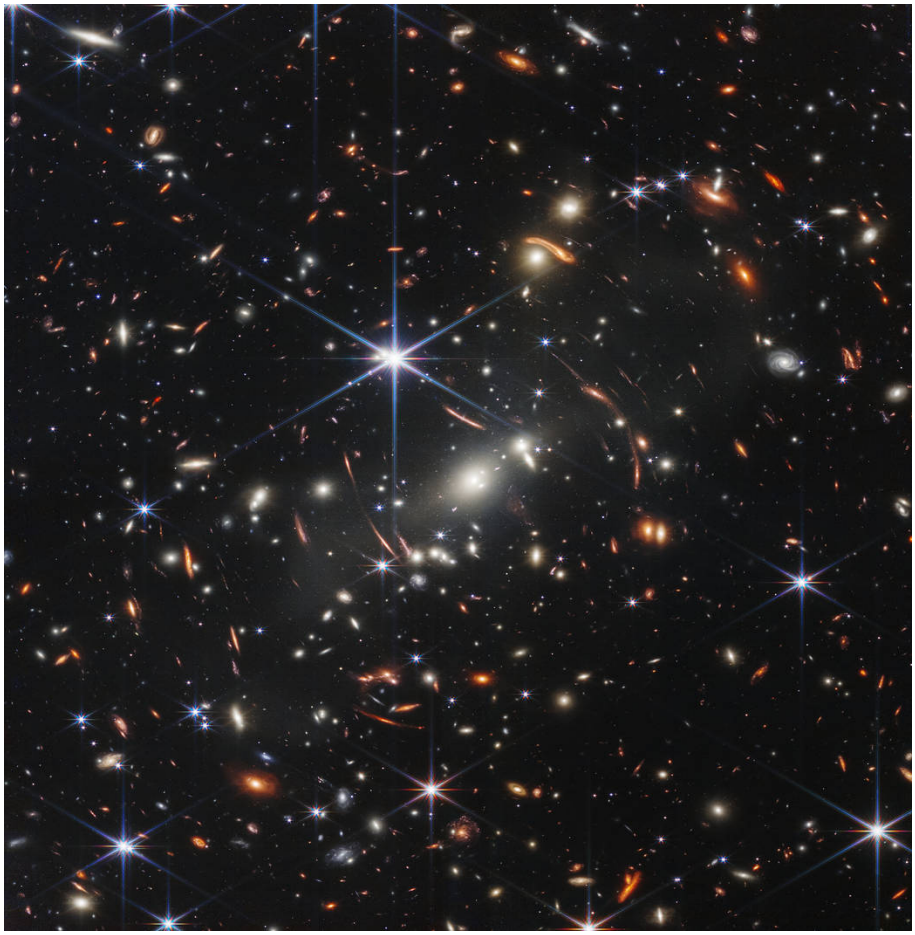


Figure 1.1. – An image of the galaxy cluster SMACS 0723 taken by the James Webb Space Telescope (and presented in Pontoppidan et al., 2022), covering an area of $2.2 \times 2.2\,\mathrm{arcmin}^2$ (Rieke et al., 2005). Aside from the massive foreground galaxy cluster and galaxies strongly distorted by the gravitational lensing effect, one can detect thousands of tiny background galaxies. Credit: NASA, ESA, CSA, STScI.

galaxies to infer the matter distribution at various redshifts via the weak gravitational lensing effect, .

To quantify cosmic shear, we analyse the shapes of millions to billions of galaxies. As the tidal gravitational fields caused by the massive structures in the Universe distort the light-rays that travel from the distant galaxy towards us, all of these shapes have been slightly deformed. To quantify this effect, a variety of two-point statistics, such as shear two-point correlation functions (2pcf, Kaiser, 1992) or Complete Orthogonal Sets of E- and B-mode Integrals (COSEBIs, Schneider et al., 2010) have been established (Kilbinger et al., 2013; Heymans et al., 2013; Joudaki et al., 2017; Troxel et al., 2018; Hikage et al., 2019; Hildebrandt et al., 2020; Wright et al., 2020; Asgari et al., 2020; Hamana et al., 2020).

As the initial density perturbations in our Universe are almost perfectly Gaussian (Planck Collaboration et al., 2020b), two-point statistics capture a large amount of the

information residing in the large-scale structure (LSS) of the Universe. However, on smaller scales, non-linear structure formation has introduced non-Gaussianities into the LSS, which two-point statistics can not capture. To extract that additional information content, a large variety of higher-order statistics has been suggested and employed to various degrees (Zürcher et al., 2021): From peak count statistics (Liu et al., 2015b,a; Kacprzak et al., 2016; Martinet et al., 2018; Harnois-Déraps et al., 2021), density split statistics (Gruen et al., 2018; Burger et al., 2020, 2022), shear clipping (Giblin et al., 2018) and Minkowski functionals (Petri et al., 2015; Marques et al., 2019; Parroni et al., 2020) to Bayesian hierarchical forward modelling (Porqueres et al., 2021, 2022), likelihood-free inference (Jeffrey et al., 2021) and the scattering transform (Cheng et al., 2020). Most of these statistics rely on *simulation-based inference*, meaning that there is no way to model them directly; instead, they are interpolated from measurements in numerical simulations.

Conversely, third-order statistics form the natural extension of the commonly used second-order statistics and can be modelled directly from the matter bispectrum. They promise to increase our constraints on cosmological parameters significantly (Kilbinger and Schneider, 2005; Kayo et al., 2013; Sato and Nishimichi, 2013). More importantly, they enable us to constrain other astrophysical or observational effects, like how galaxies align with their surrounding tidal gravitational field, without additional observations or simulations (Pyne and Joachimi, 2021; Pyne et al., 2022). Not only does this allow us to test our understanding of these effects, but it also further increases our potential constraining power. So far, third-order shear statistics have been used exactly once to constrain cosmological parameters in a weak lensing survey (Fu et al., 2014), although there appear to be efforts currently directed towards such an analysis on more recent data (Secco et al., 2022a). In galaxy clustering studies, third-order statistics are a bit more widespread and have been used to detect signs of a deviation of the standard interpretation of dark energy (Solà Peracaula et al., 2019).

In this work, we[1] will introduce persistent homology as a tool for simulation-based inference of cosmological parameters. Persistent homology is a tool from topological data analysis, which has already seen some applications in cosmology, for example, to study the topology of the cosmic web (Sousbie, 2011; van de Weygaert et al., 2013), magnetic fields (Makarenko et al., 2018), the reionisation bubble network (Elbers and van de Weygaert, 2019) and the cosmic microwave background (Pranav et al., 2017, 2019). Furthermore, Xu et al. (2019) developed an effective void finder based on persistent homology. We were the first to quantify the constraining power of persistent homology for cosmological parameters in Heydenreich et al. (2021). Soon after, Biagetti et al. (2020, 2022) published a similar study utilising galaxy clustering instead of cosmic shear. Additionally, we performed the only cosmological parameter analysis with persistent homology to date (Heydenreich et al., 2022a).

Afterwards, we will present our efforts to prepare for a cosmological parameter analysis with third-order shear statistics. While such an analysis has been performed before on data of Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) by Fu et al. (2014), the current generation of cosmic shear surveys like the Kilo-Degree Survey (KiDS, de Jong et al., 2013) or Dark Energy Survey (DES, Abbott et al., 2018) span 10 to 40 times the area

[1]Unless I want to explicitly highlight the contributions from individual people, I will use the scientific wording 'we' throughout this thesis.

(Jee et al., 2013), meaning that the impact of modelling uncertainties and astrophysical systematics is significantly more severe. If unaccounted for, these effects would bias the cosmological parameter constraints so much that the final constraints would be practically meaningless (Semboloni et al., 2008). In this work, we will present our modelling pipeline for third-order shear statistics, which we have validated and tested in numerous ways to ensure that we are aware of its specific limitations (Heydenreich et al., 2022b). We will further present our efforts to model a covariance for third-order shear statistics (Linke et al. in prep.). Once a model for the most relevant systematic biases has been found, we can perform a cosmological parameter analysis of current-generation surveys such as KiDS or DES.

This thesis is organised as follows: In Chap. 2, we will present the theoretical background necessary to understand this thesis. We will start by summarising the standard model of cosmology and the weak gravitational lensing formalism and then give a brief overview of some mathematical and statistical tools central to this thesis. We will briefly introduce the data products used in this thesis in Chap. 3. In Chap. 4, we will present our efforts to constrain cosmological parameters using persistent homology. Afterwards, in Chap. 5, we will detail our modelling and validation efforts for third-order shear statistics. We will present some concluding remarks in Chap. 6 and give an outlook on future work.

CHAPTER 2

Theoretical background knowledge

"Aristotle said a bunch of stuff that was wrong. Galileo and Newton fixed things up. Then Einstein broke everything again. Now, we've basically got it all worked out, except for small stuff, big stuff, hot stuff, cold stuff, fast stuff, heavy stuff, dark stuff, turbulence, and the concept of time."
– Zach Weinersmith

This chapter briefly summarises the background knowledge necessary to understand the thesis. In the first section, we introduce the standard model of cosmology, the ΛCDM model. The second section focuses on the theory behind the gravitational lensing effect. In the third section, we review some mathematical and statistical methods required for this thesis.

2.1. Cosmology

In its essence, the field of cosmology studies the origin and evolution of the Universe as a whole (Ellis, 2006). By studying a large variety of observables, like the CMB or the LSS of matter, we investigate how the Universe began, how it developed, and which laws of nature govern its evolution on cosmological distance- and time-scales. We summarise our present knowledge about cosmology while focusing on the aspects relevant to the work conducted in this thesis. Our summary is based on Peebles (2020); Dodelson and Schmidt (2020); Schneider (2006); Schneider (2009); the interested reader can find more detailed explanations there.

2.1.1. The standard model of cosmology

In this section, we give a brief overview of the standard model of cosmology and a few potential extensions, which we take as a basis for all analyses performed in this thesis. The theoretical foundation of this model is the revolutionary theory of general relativity (GR)

discovered by Einstein (1915), where he postulated that our three spatial dimensions are entangled with time and form a four-dimensional spacetime manifold. The geometry of this manifold is dictated by the presence of matter and energy via the field equations

$$R_{\mu\nu} - \frac{R}{2}\, g_{\mu\nu} + \Lambda\, g_{\mu\nu} = \frac{8\pi G}{c^4}\, T_{\mu\nu} \ . \tag{2.1}$$

Here, $R_{\mu\nu}$ is the Ricci curvature tensor and R the scalar curvature, both of which can be derived from the metric tensor $g_{\mu\nu}$. These quantities describing spacetime geometry depend on the energy-momentum tensor $T_{\mu\nu}$. The term Λ was originally introduced by Einstein to allow for a static Universe; now, it describes the effects of dark energy. Within the framework of GR, all free-falling particles (including photons) move through spacetime on *geodesics*, which represent straight lines following the manifold's curvature. In other words, matter dictates how spacetime curves and that curvature, in turn, dictates how matter moves. To describe the motion of matter in our Universe, we need to find its *metric*.

2.1.1.1. The Friedmann–Lemaître–Robertson–Walker metric

Solving Einstein's field equations is notoriously difficult and can only be done under some simplifying assumptions. To derive the metric of our Universe, we rely on the *cosmological principle*, which states that on large scales ($\gtrsim 200\,\mathrm{Mpc}$), the Universe is homogeneous and isotropic. The isotropy of the Universe can be observed to a very high degree both in the large-scale structure of the galaxy distribution (Eisenstein et al., 2011) and in the temperature of the CMB (Migliaccio and Planck Collaboration, 2019). In recent years, some doubts have been cast upon this assumption of isotropy: There are claims that the fine-structure constant (Wilczynska et al., 2020), the Hubble constant (Luongo et al., 2022) and the temperature-luminosity relation of galaxy clusters (Migkas et al., 2020, 2021) are not isotropic. While it is still too early to rule out observational effects, the assumption of isotropy may have to be abandoned in the future.

Homogeneity, on the other hand, is much more challenging to observe, as we can not observe our Universe at a uniform time due to the finite speed of light. Instead, we have to assume that the Universe is isotropic for all observers, meaning that our place in the Universe is not unique in any way. From this, homogeneity follows quickly.

Given a homogeneous and isotropic Universe, we can construct the Friedmann–Lemaître–Robertson–Walker (FLRW) metric as (Robertson, 1935; Walker, 1937)

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left[\mathrm{d}\chi^2 + f_K^2(\chi)\,(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)\right] \ . \tag{2.2}$$

Here, t is the cosmic time, θ and φ are spherical angular coordinates. χ denotes the *radial comoving coordinate* and $a(t)$ is the *scale factor*, describing the relative size of the Universe, normalised to $a_0 = a(t_0) = 1$ today[1]. At last, $f_K(\chi)$ is the comoving angular diameter distance, depending on the curvature of the Universe via

$$f_K(\chi) = \frac{1}{\sqrt{K}} \sin\left(\sqrt{K}\chi\right) \ , \tag{2.3}$$

[1]Here, and throughout this thesis, we denote the value of cosmological variables today by the subscript $_0$.

where complex notation accounts for the case of negative curvature. In particular, $f_K(\chi) = \chi$ holds for a flat Universe with $K = 0$. A universe whose spacetime is subject to the FLRW metric is not necessarily static; it can contract or expand (in fact, ours does the latter). Therefore, it makes sense to introduce *comoving coordinates* $\boldsymbol{x}$ as

$$\boldsymbol{r}(t) = a(t)\boldsymbol{x} \,, \tag{2.4}$$

where $\boldsymbol{r}(t)$ is the proper (physical) coordinate. Two observers with no peculiar velocities always have the same comoving distance in an expanding spacetime. If we can determine the history of the scale factor $a(t)$, we can infer the dynamics of the Universe on the largest scales.

In order to do that, we insert Eq. (2.2) into the field equations (2.1) and assume that the energy-momentum tensor describes a perfect fluid with density ρ and pressure p. From this, the *Friedmann equations* (Friedmann, 1922) can be derived, which describe the temporal behaviour of the Universe's expansion rate:

$$H^2(a) \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3} \,, \tag{2.5}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \,, \tag{2.6}$$

where G is the gravitational constant. The first Friedmann equation (2.5) describes the expansion rate $H(a)$ of the Universe, with the expansion rate today, $H_0 = H(t_0)$, called the Hubble constant. After its first measurement by Hubble (1929a), which was off by a factor of 8, this Hubble constant has been measured to almost per cent-level precision. The curious part is that observations of the early Universe like BAOs and the CMB suggest a value of $H_0 \approx 68\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Planck Collaboration et al., 2020b), whereas observations of the local Universe like supernovae of type Ia (Riess et al., 2022), masers (Pesce et al., 2020), surface brightness fluctuations (Blakeslee et al., 2021) or strong gravitational lensing time delays (Wong et al., 2020) suggest a value of $H_0 \approx 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The exact amount of tension varies depending on the measurements taken into account but is around the 5σ level. Due to the residual uncertainty in the Hubble constant and the fact that many other parameters directly depend on it, we usually parametrise it as

$$H_0 = 100\,h\frac{\mathrm{km}}{\mathrm{s\,Mpc}} \,, \tag{2.7}$$

where h is the dimensionless Hubble constant.

The Hubble expansion gives rise to peculiar phenomena. For photons, which describe null-geodesics with $\mathrm{d}s^2 = 0$ in the spacetime metric (2.2), $c^2\mathrm{d}t^2 = a^2(t)\mathrm{d}\chi^2$ holds. In particular, when we observe a distant source, observed time-intervals Δt_obs and thus frequencies ν_obs and wavelengths λ_obs differ from their values at the time they were emitted,

$$\frac{a(t_\mathrm{obs})}{a(t_\mathrm{em})} = \frac{1}{a(t_\mathrm{em})} = \frac{\nu_\mathrm{em}}{\nu_\mathrm{obs}} = \frac{\lambda_\mathrm{obs}}{\lambda_\mathrm{em}} \,. \tag{2.8}$$

We define the *redshift* z of an observed source as the ratio of the wavelengths,

$$1 + z = \frac{\lambda_\mathrm{obs}}{\lambda_\mathrm{em}} = \frac{1}{a_\mathrm{em}} \,. \tag{2.9}$$

Locally, this redshift is often interpreted as a Doppler-shift introduced by the recession velocity of the observed galaxy,

$$v = H_0 D = cz \ , \tag{2.10}$$

but this picture is certainly not correct on cosmological distance scales, as we have observed galaxies with redshift $z \gg 1$, which would indicate a recession velocity $v \gg c$ (for a more thorough discussion, see Davis and Lineweaver, 2004).

Combining both Friedmann equations, we get the adiabatic equation

$$\frac{\mathrm{d}(a^3 \rho c^2)}{\mathrm{d}t} = -p \frac{\mathrm{d}a^3}{\mathrm{d}t} \ , \tag{2.11}$$

which dictates that the energy density ρc^2 within a comoving volume changes with the work done by its expansion or contraction, in a cosmological equivalent to energy conservation. From this equation, we can infer how the density of different matter components changes with time, given their equation of state (EoS)

$$w = \frac{p}{\rho c^2} \ . \tag{2.12}$$

In an ideal fluid, particles whose energy is dominated by their momentum have an EoS of $w = 1/3$; in our present Universe, they are mainly photons and neutrinos and contribute to the *radiation density* ρ_r. Particles whose energy is dominated by their rest-mass are pressureless with an EoS of $w = 0$ and contribute to the *matter density* ρ_m in our Universe. There, we distinguish baryonic matter, which is made up of protons, neutrons, and electrons, from dark matter, a mysterious collisionless matter component that appears to interact only gravitationally. In the ΛCDM model, we assume the presence of dark energy, which has an EoS of $w = -1$. This corresponds either to a constant term in the field equations of GR (the Λ in Eq. 2.1) and thus a property of gravity, or to a fluid with constant density $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$ and negative pressure, the *vacuum energy*. This would correspond to writing the term $\frac{\rho_\Lambda 8\pi G}{c^2}$ on the right side of Eq. (2.1). Unfortunately, both interpretations result in the same expansion behaviour of the Universe and can thus not be distinguished using cosmological observations. For practical purposes, we use the vacuum energy interpretation of dark energy for this thesis.

Of course, our knowledge of dark energy is extremely limited, and many interpretations of this phenomenon exist. To constrain different dark energy models, we often extend the standard ΛCDM model by allowing the EoS of dark energy to be different, $w \neq -1$, which we call the w cold dark matter model (wCDM) model. Sometimes, we even invoke dark energy with an EoS that varies with time, $w(a) = w_0 + (1-a)w_a$, the $w_0 w_a$ cold dark matter model ($w_0 w_a$CDM) model.

Solving the adiabatic equation (2.11) for an arbitrary fluid with density ρ and EoS $w(a)$, we get

$$\rho(a) = \rho_0 \exp\left[3 \int_a^1 \mathrm{d}a' \, \frac{1 + w(a')}{a'}\right] \ . \tag{2.13}$$

When we assume an EoS that evolves linearly with the scale factor, $w(a) = w_0 + (1-a)w_a$, we solve the equation above to

$$\rho(a) = \rho_0 \left[a^{-3(1+w_0+w_a)} \mathrm{e}^{-3w_a(1-a)}\right] \ . \tag{2.14}$$

In particular, we can insert the EoS for radiation, matter and dark energy:

$$\begin{array}{rll} \text{matter} & w_\mathrm{m} = 0 & \rho_\mathrm{m}(a) \propto a^{-3} \\ \text{radiation} & w_\mathrm{r} = \frac{1}{3} & \rho_\mathrm{r}(a) \propto a^{-4} \\ \text{dark energy} & w_\Lambda = -1 & \rho_\Lambda(a) = \text{const.} \end{array}$$

To describe the evolution of the Universe, we define the critical density $\rho_{\mathrm{crit},0} = \frac{3H_0^2}{8\pi G}$, which is the energy density that the Universe today has to assume such that the curvature term K vanishes in the first Friedmann equation (2.5). We note that due to our interpretation of dark energy as vacuum energy, the Λ-term in Eqs. (2.5) and (2.6) vanishes, and the density ρ takes the form $\rho = \rho_\mathrm{m} + \rho_\mathrm{r} + \rho_\Lambda$. Using this, we can define the dimensionless density parameters $\Omega_x = \rho_{x,0}/\rho_{\mathrm{crit},0}$ and rewrite the Friedmann equation as

$$H^2(a) = \left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \left[\Omega_\mathrm{r}\, a^{-4} + \Omega_\mathrm{m}\, a^{-3} + (1-\Omega_0)\, a^{-2} + \Omega_\Lambda\right] , \tag{2.15}$$

where we have defined $1 - \Omega_0 = 1 - \Omega_\mathrm{m} - \Omega_\mathrm{r} - \Omega_\Lambda = Kc^2/H_0^2$. If we can determine the values of these density parameters today, we can describe the Universe's expansion behaviour as a solution to this differential equation.[2] Therefore, determining the values of these cosmological parameters is of great interest. The *Planck* mission determined those parameters with great precision: We know that our Universe is almost completely flat with $|1 - \Omega_0| < 0.003$ and that the matter density is about $\Omega_\mathrm{m} \approx 0.315$ (Planck Collaboration et al., 2020b). The radiation density parameter can be calculated easily from the CMB-Temperature and yields $\Omega_\mathrm{r} = \Omega_\gamma + \Omega_\nu \approx 1.68\,\Omega_\gamma \approx 4.2\,h^{-2} \times 10^{-5}$, where Ω_γ is the photon energy density parameter and Ω_ν the neutrino energy density parameter.

We can use the definition of the Hubble parameter $H(a) = \dot{a}/a$ and the FLRW metric to derive some useful relations between time, scale factor, comoving distance, and redshift:

$$\mathrm{d}a = -\frac{1}{(1+z)^2}\mathrm{d}z, \quad \mathrm{d}t = \frac{\mathrm{d}a}{aH(a)} = -\frac{\mathrm{d}z}{(1+z)H(z)}, \quad \mathrm{d}\chi = -\frac{c\,\mathrm{d}t}{a} = -\frac{c\,\mathrm{d}a}{a^2 H(a)} = \frac{c\,\mathrm{d}z}{H(z)} . \tag{2.16}$$

Using the relation between time t and scale factor a, we can determine how the relative size of the Universe evolves with time. Unfortunately, there is no analytic solution for a Universe containing matter, radiation and dark energy. However, we know that the expansion behaviour of the Universe was governed by its different components at different times, as depicted in Fig. 2.1. In the very early Universe, radiation was the dominating component (as $\rho_\mathrm{r} \propto a^{-4}$). After that followed a period of matter domination. Only recently, our Universe has begun to be dominated by dark energy. The fact that we live in a time where a transition between the dominance of two components (matter and dark energy) is taking place seems like another unlikely coincidence. One can not help but wonder if there are other components, whose densities are just extremely low today, but that either dominated the energy budget of the Universe at very early times or will dominate it in the future. Assuming the (flat) Universe contains only a single component x with EoS w (which is a good approximation for many cosmological epochs), we can write the Hubble function as

$$H^2(a) = H_0^2\, a^{-3(1+w)} . \tag{2.17}$$

[2] Of course that only holds if the ΛCDM model is accurate. For example, the future expansion of the Universe depends massively on the EoS of dark energy, compare Fig. 2.1.

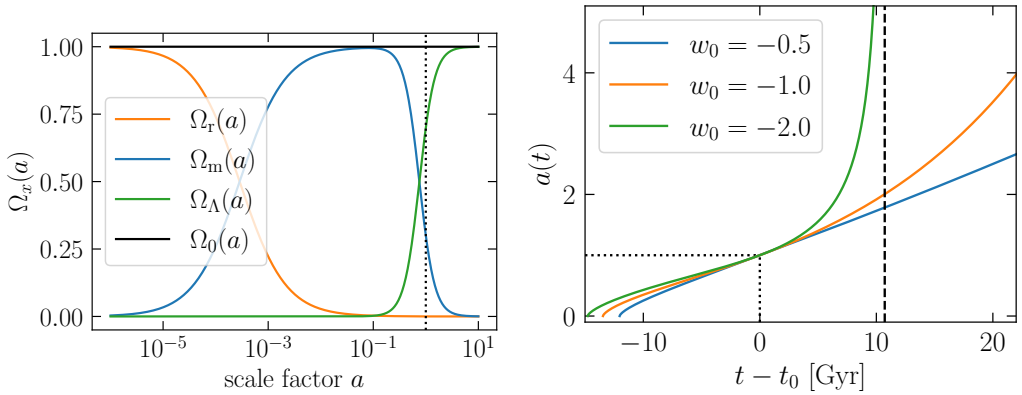


Figure 2.1. – The left panel shows the value of the matter density parameters in a flat ΛCDM Universe with $H_0 = 70\,\frac{\mathrm{km}}{\mathrm{s\,Mpc}}$ and $\Omega_\mathrm{m} = 0.3$ as a function of the scale factor. The right panel shows the behaviour of the scale factor as a function of time for three flat wCDM cosmologies with $H_0 = 70\,\frac{\mathrm{km}}{\mathrm{s\,Mpc}}$ and $\Omega_\mathrm{m} = 0.3$. One can see that the scale factor a diverges at $t \approx 10.7\,\mathrm{Gyr}$ for the case of $w = -2$ (shown by the black dashed line). In both cases, the dotted lines indicate the value of the scale factor a and cosmic time t today. Adapted from Jens Erler.

Inserting this into Eq. (2.16), we get

$$t(a) = \begin{cases} \frac{2}{3H_0(1+w)} a^{\frac{3}{2}(1+w)} & w \neq -1 \\ \frac{1}{H_0} \ln(a) + t_0 & w = -1 \end{cases}, \tag{2.18}$$

yielding

$$a(t) = \begin{cases} \left[\frac{3H_0(1+w)}{2} t\right]^{\frac{2}{3}(1+w)^{-1}} & w \neq -1 \\ \exp(H_0(t - t_0)) & w = -1 \end{cases}. \tag{2.19}$$

Note that for the case of $w < -1$, we have been a bit imprecise with our notation, as we implicitly assumed that at $t = 0$, $a = 0$ holds. However, for $w \leq -1$ the Universe never has a scale factor of $a = 0$, which implies that a Universe containing only dark energy would have no big bang. We can see that in a Universe that after some time becomes dominated by a component with EoS $w < -1$, the scale factor $a(t)$ diverges at some finite time t, which would imply that the distances between every single elementary particle in the Universe become infinite. This is called the 'big rip' scenario. As an example, we show the expansion behaviour of the Universe for different dark energy EoS in a wCDM cosmology in Fig. 2.1.

We can now insert the already known EoS for matter, radiation and dark energy into Eq. (2.19), yielding $a(t) \propto t^{\frac{2}{3}}$ for a matter-dominated Universe, $a(t) \propto t^{\frac{1}{2}}$ for a radiation-dominated Universe and $a(t) \propto \exp(H_0 t)$ for a Universe dominated by dark energy with an EoS of $w = -1$. The current age of the Universe, t_0, can be measured both by constraining the other cosmological parameters (Planck Collaboration et al., 2020b) and by investigating the age of the oldest objects in our Universe (Valcin et al., 2020). Both methods are roughly consistent (Di Valentino et al., 2021c) and yield $t_0 \approx 13.8\,\mathrm{Gyr}$.

2.1.1.2. Cosmological distances

Living in an expanding Universe with a finite speed of light and a non-Euclidean metric gives rise to many phenomena that seem counter-intuitive to our minds, which have adjusted to a static, Euclidean spacetime. For example, we know that looking far away means looking into the past and that we look towards the singularity at the big bang in every direction. We can further show that 94% of all galaxies we can observe are no longer in our future light cone (meaning that we can never reach them, even if we travel at the speed of light towards them). While these phenomena are relatively easy to understand, they contradict our intuition.

In cosmology, the probably most relevant notion is that the distance to an object can no longer be uniquely defined. Different concepts of distance have emerged on Earth, which are all equivalent in our minds, yet yield vastly different results when applied on cosmological scales. For example, to get the distance to an object, we could divide its (known) physical size R by its angular extent $\mathrm{d}\theta$ and get the *angular diameter distance* $D_{\mathrm{ang}} = R/\mathrm{d}\theta$. Alternatively, if the object radiates energy with a luminosity L, we can use the inverse square law and calculate the distance from the measured flux S via $D_{\mathrm{lum}} = \sqrt{L/(4\pi S)}$. We could also measure the time τ it takes a light-ray[3] to travel that distance and get the lookback distance $D_\gamma = c\tau$. On Earth, all these measures will yield the same result (barring measurement errors). In an expanding Universe, these distances are vastly different, which gives rise to some confusing stories (especially in non-scientific media). For example, when the Hubble Space Telescope (HST) detected the furthest star ever recorded (Welch et al., 2022), it was reported by *ctvnews* to be more than 28 billion light-years away[4], which seems odd as our Universe is only about 14 billion years old. Other sources, like the *independent*[5], cite the distance to the same star as 13 billion light-years. Despite these vastly different distance estimates, none of the two newspapers made a mistake; they simply used different distance measures.

From the FLRW metric (2.2), we can set $\mathrm{d}s = R$ and get

$$D_{\mathrm{ang}}(z) = \frac{R}{\mathrm{d}\theta} = \frac{a f_K(\chi)\mathrm{d}\theta}{\mathrm{d}\theta} = \frac{f_K(\chi)}{1+z} \,. \tag{2.20}$$

More generally, defining $D_{\mathrm{ang}}(z_1, z_2)$ and $\chi(z_1, z_2)$ as the angular diameter and comoving distances between redshifts z_1 and z_2, we can calculate

$$D_{\mathrm{ang}}(z_1, z_2) = \frac{f_K[\chi(z_1, z_2)]}{1+z_2} \,. \tag{2.21}$$

Note that, while $\chi(z_1, z_2) = \chi(z_2) - \chi(z_1)$ holds, $f_K(z_1, z_2) = f_K(z_2) - f_K(z_1)$ only holds in a flat universe, and $D_{\mathrm{ang}}(z_1, z_2) \neq D_{\mathrm{ang}}(z_2) - D_{\mathrm{ang}}(z_1)$ in general.

A similar relation for the luminosity distance can be obtained (Etherington, 1933):

$$D_{\mathrm{lum}}(z) = \sqrt{\frac{L}{4\pi S}} = (1+z) f_K(\chi) = (1+z)^2 D_{\mathrm{ang}}(z) \,. \tag{2.22}$$

[3] Like all physicists, we assume that we live in a perfect vacuum.

[4] `https://www.ctvnews.ca/sci-tech/hubble-sees-most-distant-star-ever-28-billion-light-years-away-1.5840991`

[5] `https://www.independent.co.uk/news/uk/earth-scientists-relics-johns-hopkins-university-albert-einstein-b2047376.html`

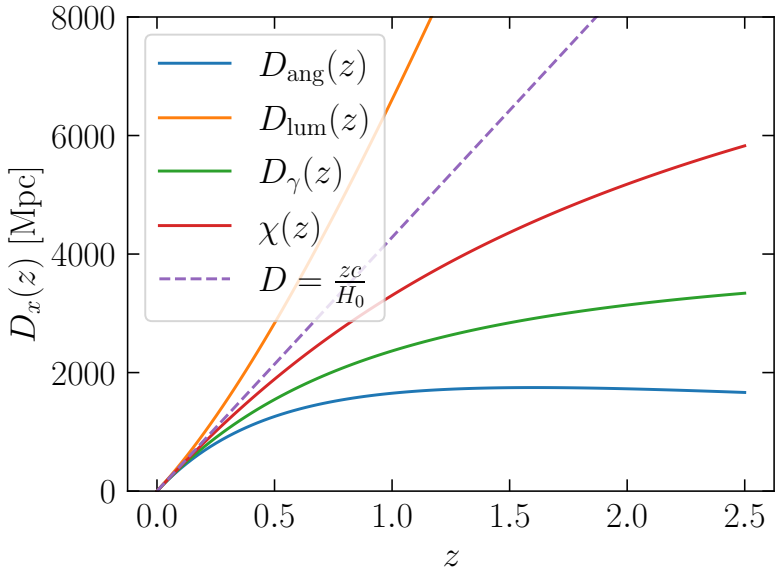


Figure 2.2. – The different distance measures as a function of redshift z in a flat ΛCDM cosmology. We show the angular diameter distance D_{ang} (blue), the luminosity distance D_{lum} (orange), the lookback distance D_γ (green), the comoving distance χ (red) and the linear Hubble law $D = zc/H_0$ (purple, dashed). One can see that for small redshifts $z \lesssim 0.1$, the distance measures agree.

We can further show with Eq. (2.16), that

$$D_\gamma(z) = c\tau = c\int_{t_{\mathrm{em}}}^{t_{\mathrm{obs}}} \mathrm{d}t = \int_0^z \frac{\mathrm{d}z'\, c}{(1+z')H(z')}\,, \tag{2.23}$$

holds.

As a little example, we can take the star observed by Welch et al. (2022), which has a redshift of $z = 6.2$. Assuming a flat universe with $H_0 = 68\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_{\mathrm{m}} = 0.3$, we get

$$f_K(\chi) = \chi \approx 28.0\times10^9\,\mathrm{ly},\ D_\gamma \approx 13.0\times10^9\,\mathrm{ly},\ D_{\mathrm{ang}} \approx 3.9\times10^9\,\mathrm{ly},\ D_{\mathrm{lum}} \approx 201.3\times10^9\,\mathrm{ly}\,. \tag{2.24}$$

We can see that *ctvnews* cited the comoving distance to the star, while *independent* opted for the lookback distance.

A graph of all distance measures as a function of redshift is shown in Fig. 2.2. It is noteworthy that the angular diameter distance D_{ang} has a maximum at $z \approx 1.7$ and decreases afterwards.

2.1.1.3. Cosmic inflation

Within the ΛCDM model, we can now compute the Universe's expansion history. When doing this, two observations stick out that are hard to explain.

The first one is the *flatness problem* (Dicke, 1970): We observe our Universe to have a very low curvature if any. This means that the sum of all energy densities is almost equal

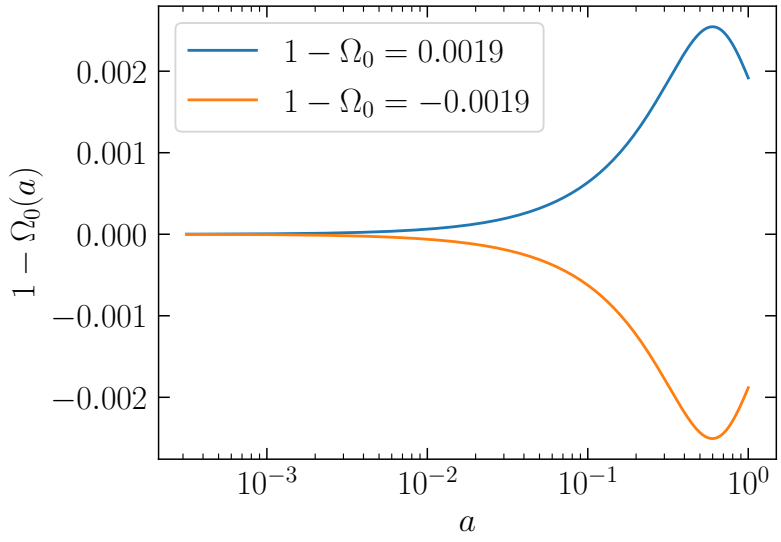


Figure 2.3. – Two examples for the evolution of curvature with the scale factor a in a Universe with $\Omega_\mathrm{m} = 0.3$ and $\Omega_\Lambda \approx 0.7$.

to the critical density of the Universe (nowadays, we know that with an accuracy of about 0.2%, see Planck Collaboration et al., 2020b). If we track the evolution of curvature via

$$1 - \Omega_0(a) = 1 - \frac{\rho_\mathrm{m}(a) + \rho_\mathrm{r}(a) + \rho_\Lambda(a)}{\rho_\mathrm{crit}(a)} \,, \tag{2.25}$$

we see that any potential curvature today has been much smaller at early times. In particular, at $t \approx 10^{-34}\,\mathrm{s}$, the curvature parameter must have been smaller than 10^{-52}. Such an incredible coincidence seems extremely unlikely, meaning there has to be a mechanism in the early Universe to drive down the curvature.[6]

The second problem is the *horizon problem* (Rindler, 1956). It stems from the observation that the CMB around us is extremely isotropic to a degree of $\Delta T/T \lesssim 10^{-5}$. However, we can show that large parts of the CMB have not been in causal contact with each other, meaning that they could not have exchanged any information since the birth of the Universe: The distance that light (and thus information) can travel in an expanding Universe can be parametrised by the comoving horizon

$$r_\mathrm{H,com} = \int_0^t \frac{c\,\mathrm{d}t}{a(t)} = \int_z^\infty \frac{c\,\mathrm{d}z}{H(z)} \,. \tag{2.26}$$

The CMB was released at a redshift of $z \approx 1100$, giving a comoving horizon size of $r_\mathrm{H,com} \approx 280\mathrm{Mpc}$. To compute its angular extent on the sky, we need to employ the comoving angular diameter distance $f_K(\chi(z)) = \chi(z)$, where we assume a flat Universe.

[6] Some people rely here on the *anthropological argument*, which states that a too large curvature would not allow for a Universe that has active star formation for a sufficient amount of time to form life, meaning that we live in one of the few Universes that allow for the existence of observers. However, that argument requires the existence of infinite universes and seems quite unsatisfactory.

This can be calculated from Eq. (2.16) and takes $\chi(1100) \approx 13.9\text{Gpc}$. The angular extent of a causally connected region of the CMB on the sky is thus

$$\theta = \frac{d_{\text{H,com}}}{\chi} \approx 1^\circ 9' \,. \tag{2.27}$$

This means that, without ever having any chance of exchanging information, different regions of the CMB have incredibly similar temperatures. Even more puzzling, the tiny temperature fluctuations of the CMB are correlated on scales beyond the horizon size without an opportunity to exchange information (Planck Collaboration et al., 2020b; Cea, 2022). Again, this seems like too big of a coincidence to ignore.

A proposed solution for these two problems is called *inflation*. The basic principle is the following: The Grand Unified Theories (GUT) propose that at a temperature of $T \sim 10^{14}\,\text{GeV}$, the electromagnetic force and both strong and weak nuclear forces arise from a single gauge theory due to symmetry breaking. We have no way to create these settings in a laboratory, so we have no idea what happens at those temperatures. However, there is a good reason to assume that something beyond our standard models occurs. These temperatures were exceeded in the early Universe until about $t \approx 10^{-34}\,\text{s}$.

We propose that before this time, the Universe was subject to an exponential expansion (similar to present-day dark energy, just 27 orders of magnitude stronger), driven by a slow-rolling scalar field. In this rapidly expanding Universe, particle-antiparticle pairs would spontaneously emerge and immediately be separated by the rapid expansion. Fluctuations in the expansion caused by the energy-time uncertainty would lead to tiny perturbations in the extraordinarily homogeneous Universe and form the primordial seeds of structure formation.[7] This scenario would have several consequences:

As can be seen in Fig. 2.3, an accelerated expansion decreases the curvature of the Universe. After about 60 e-foldings, any curvature of order unity is smoothed out to the level we observe in our Universe. Inflation would thus naturally solve the flatness problem. Furthermore, before inflation occurred, the entire observable Universe was in causal contact and thus could come to its homogeneous state. This would solve the horizon problem.

Additionally, this phenomenon would solve a different issue that GUT face. At extremely high temperatures, these theories predict the emergence of heavy and exotic particles, like magnetic monopoles, which should dominate over baryonic and dark matter. In an inflationary scenario, the density of these exotic particles would be decreased to a negligible amount, which would explain why we have not found them yet.[8]

The major success of inflation is that it predicts the slope of the primordial power spectrum n_s (which we introduce in Sect. 2.1.2.5) to be almost unity but a bit smaller. Our best constraints on this parameter are $n_s = 0.965 \pm 0.004$, which confirm this prediction.

Whether or not inflation actually happened, and if so, how, is still actively debated. Some people claim that the conditions required for this inflationary process require even more fine-tuning than the flatness problem (Penrose, 2016; Steinhardt, 2011), and current-

[7] The question of why matter heavily dominates over antimatter in the current Universe is called *baryogenesis* and has not yet been answered satisfactorily.

[8] On the other hand, to explain the non-existence of theorised particles with a theorised scenario seems a bit pointless.

generation observations have already ruled out many models of inflation. Others argue that current observations support the scenario of inflation (Guth et al., 2014).

Unfortunately, constraining inflation is challenging, as we can not directly observe it. In the future, we hope to observe B-mode polarisation patterns in the CMB[9], which would be strong evidence for inflation and could place meaningful constraints on different inflation scenarios. Furthermore, the planned Laser Interferometer Space Antenna (LISA) may be able to measure a background of primordial gravitational waves, which would have been caused by inflation.

2.1.2. Structure evolution in the Universe

By now, we have discussed the evolution of the Universe as a whole and found a way to describe its expansion rate with a relatively simple equation. We now want to direct our attention to the structures that form within this expanding Universe.

2.1.2.1. The origin of structure

As we can observe from the CMB, the early Universe was extremely homogeneous with tiny anisotropies of $\frac{\Delta T}{T} \sim 10^{-5}$. In this work, we assume the widely accepted theory that the seeds of structure formation are tiny quantum fluctuations that have been blown up to macroscopic scales by the inflation of the early Universe. These seeds give rise to tiny fluctuations in the gravitational potential Φ, so overdensities attract more matter and grow over time. At some point, they become dense enough to collapse into gravitationally bound dark matter halos. These dark matter halos then begin to merge with others and accrete into galaxies and, eventually, massive galaxy clusters. In this section, we want to quantify the growth of these matter perturbations over time and find a model for their eventual collapse.

It is noteworthy that before the CMB was released, the baryonic matter formed a charged plasma, consisting primarily of protons, helium nuclei and electrons. This plasma was strongly coupled to the radiation via Compton scattering. While the gravitational potential attracted the baryons, the radiation pressure suppressed the infall of baryons into overdensities. This gave rise to oscillatory modes in the baryon-photon fluid. Once the Universe cooled down enough so that the baryons could form neutral elements, it became transparent to photons. The CMB was released, and the baryons could fall into the overdensities that the dark matter distribution had formed. The BAOs of the baryon-photon fluid are imprinted both in the CMB and in the LSS of galaxies and contain a wealth of cosmological information. However, for our work, they are not relevant, so we focus on the structure evolution of dark matter, which poses the dominant matter component in the Universe and has remained unaffected by the radiation pressure. The distribution of baryonic matter largely follows the dark matter distribution; however, the formation of stars and galaxies is subject to many hydrodynamical processes that are more complex than the simple gravitational interaction of dark matter. Therefore, the exact correspondence between the emergence of galaxies and the underlying dark matter distribution is more complicated.

[9] BICEP2 Collaboration et al. (2014) reported a detection of B-mode polarisation, but it was later shown that their signal might have been caused by Galactic dust

2.1.2.2. Linear evolution of matter density perturbations

To understand the growth of structure in the Universe, we employ the pressureless hydrodynamical fluid equations

$$\frac{\partial \rho}{\partial t} + \nabla_r \cdot (\rho \boldsymbol{v}) = 0 \,, \tag{2.28}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla_r) \cdot \boldsymbol{v} = -\nabla_r \Phi \,, \tag{2.29}$$

$$\nabla_r^2 \Phi = 4\pi G \rho - \Lambda \,. \tag{2.30}$$

For a fluid, the *continuity equation* (2.28) describes the conservation of mass, the *Euler equation* (2.29) corresponds to the conversation of momentum in the presence of a gravitational potential Φ. The *Poisson equation* (2.30) describes how the presence of matter shapes the gravitational potential Φ. This equation has been modified to allow for the presence of a cosmological constant Λ.

We now introduce the comoving density, velocity and gravitational potential (compare Eq. 2.4):

$$\varrho(\boldsymbol{x}, t) = \varrho[\boldsymbol{r}/a(t), t] = \rho(\boldsymbol{r}, t) \,, \tag{2.31}$$

$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{v}(\boldsymbol{r}, t) - \dot{a}\boldsymbol{x} = \boldsymbol{v}\left(\boldsymbol{r}, t\right) - \frac{\dot{a}}{a}\boldsymbol{r} \,, \tag{2.32}$$

$$\Phi(\boldsymbol{x}, t) = \Phi(a\boldsymbol{x}, t) + \frac{\ddot{a}a}{2} x^2 \,. \tag{2.33}$$

Furthermore, we define the density contrast $\delta(\boldsymbol{x}, a) = [\varrho(\boldsymbol{x}, a) - \bar{\varrho}(a)]/\bar{\varrho}(a)$ and note that $\nabla_x = a\nabla_r$. We find that a completely homogeneous universe with $\rho(\boldsymbol{x}, t) \equiv \bar{\rho}(t)$, $\boldsymbol{v}(\boldsymbol{x}, t) = H(t)\, a(t)\, \boldsymbol{x}$, and $\Phi \equiv \frac{\ddot{a}a}{2}|\boldsymbol{x}|^2$ solves the system of Equations (2.28)-(2.30). Inserting the comoving quantities and subtracting the homogeneous solution, we get

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla_x \cdot [(1+\delta)\, \boldsymbol{u}] = 0 \,, \tag{2.34}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + \frac{\dot{a}}{a}\boldsymbol{u} + \frac{1}{a}(\boldsymbol{u} \cdot \nabla_x)\, \boldsymbol{u} = -\frac{1}{a}\nabla_x \Phi \,, \tag{2.35}$$

$$\nabla_x^2 \Phi = \frac{3H_0^2 \Omega_\mathrm{m}}{2a}\delta \,, \tag{2.36}$$

where for the last equation we have used $\bar{\varrho}(a) = \Omega_\mathrm{m}\rho_\mathrm{crit}a^{-3}$. This system of non-linear differential equations has no known general solution. However, we can use the fact that in the early Universe, the perturbations are small, meaning $\delta \ll 1$. Furthermore, we can assume that the initial peculiar velocities $\boldsymbol{u}$ are also small. Disregarding all but leading-order terms in δ and $\boldsymbol{u}$, our set of equations transforms to

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla_x \cdot \boldsymbol{u} = 0 \,, \tag{2.37}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + \frac{\dot{a}}{a}\boldsymbol{u} = -\frac{1}{a}\nabla_x \Phi \,, \tag{2.38}$$

$$\nabla_x^2 \Phi = \frac{3H_0^2 \Omega_\mathrm{m}}{2a}\delta \,. \tag{2.39}$$

We can now combine these equations by taking $\frac{\partial}{\partial t}(2.37) - \frac{\nabla}{a}(2.38)$ and use Eq. (2.39) to replace the occurring $\nabla_x^2\Phi$, yielding a second-order linear ordinary differential equation (ODE) for the density contrast:

$$\frac{\partial^2\delta}{\partial t^2} + \frac{2\dot{a}}{a}\frac{\partial\delta}{\partial t} - \frac{3H_0^2\,\Omega_\mathrm{m}}{2a^3}\delta = 0\,. \tag{2.40}$$

This ODE has two linearly independent solutions, and since no derivatives with respect to $\boldsymbol{x}$ occur in this equation, the solutions factorise. We can thus write

$$\delta(\boldsymbol{x},t) = D_+(t)\Delta_+(\boldsymbol{x}) + D_-(t)\Delta_-(\boldsymbol{x})\,. \tag{2.41}$$

By using the second Friedmann equation (2.6) for the matter-dominated era, one can see that

$$\frac{\partial}{\partial t}\left[\dot{H}(t) + H^2(t)\right] = \ddot{H}(t) + 2H(t)\,\dot{H}(t) = \frac{3H_0^2\,\Omega_\mathrm{m}}{2a^3}H(t)\,, \tag{2.42}$$

meaning that the Hubble function $H(t)$ solves Eq. (2.40). As the Hubble function decreases with time, we call this solution $D_-(t) = H(t)$. The term $\Delta_-(\boldsymbol{x})$ is not relevant for structure formation since the term $D_-(t)\Delta_-(\boldsymbol{x})$ decays with time and soon becomes negligible. Given this one solution to the ODE, the second one can be found:

$$D_+(a) \propto \frac{5\Omega_\mathrm{m}}{2}\,H(a)\,H_0^2\int_0^a \frac{\mathrm{d}a'}{[a'\,H(a')]^3}\,, \tag{2.43}$$

where we fix the proportionality constant such that $D_+(a=1) = 1$. This *growth factor* describes the relative size of density fluctuations as a function of the scale factor. One can show that for $\Omega_\mathrm{m} = 1$, $D_+(a) = a$ holds.

2.1.2.3. Structure evolution in the primordial Universe

So far, we have discussed how matter perturbations grow as a function of the scale factor in a matter-dominated Universe. Furthermore, we have worked under the assumption that density perturbations are in causal contact, meaning that the horizon size from Eq. (2.26) is much larger than the physical extent λ of the density fluctuations in question. If that is not the case, we can not apply the hydrodynamical fluid equations since gravitational interaction also propagates at the speed of light.

Right after inflation, most perturbations are much larger than the horizon size. Instead of applying the fluid equations, we treat these density perturbations as little 'pocket universes' that expand with their own Hubble flow, governed by a slightly different density ρ. Up to a scale factor of $a_\mathrm{eq} = \Omega_\mathrm{r}/\Omega_\mathrm{m}$, they are dominated by radiation. While the radiation pressure does not directly affect the dark matter perturbations, it does influence the rate of expansion and acts as a source of gravity via the energy-momentum tensor $T_{\mu\nu}$.

We will not perform these calculations here and instead refer the reader to Schneider (2009). The evolution of dark matter density perturbations can be summarised to

	super-horizon	sub-horizon
radiation-dominated	$\delta(a) \propto a^2$	$\delta(a) = \text{const.}$
matter-dominated	$\delta(a) \propto a$	$\delta(a) \propto a$

After inflation, each density perturbation starts as a super-horizon perturbation and thus grows proportional to a^2. After some time, the perturbation 'enters the horizon' (more accurately: the horizon grows larger than the density perturbation) and, provided the Universe is still dominated by radiation, the perturbation stops growing. After some more time, the Universe becomes matter-dominated, and all super- and sub-horizon perturbations grow proportional to the scale factor. This means that the growth of small density perturbations is heavily suppressed during the radiation-dominated phase, whereas perturbations larger than the horizon size at matter-radiation equality are not suppressed at all. To describe this behaviour, we employ the *transfer function* T_k.

Let $\tilde{\delta}(k, a)$ be the Fourier transform of the density contrast and let k_s be a wave number small enough that a perturbation of that scale enters the horizon in the matter-dominated era (and is thus not suppressed). Then the transfer function can be defined via

$$\frac{\tilde{\delta}(k,1)}{\tilde{\delta}(k_s,1)} = T_k \frac{\tilde{\delta}(k,a_\mathrm{i})}{\tilde{\delta}(k_s,a_\mathrm{i})} \ , \tag{2.44}$$

where a_i denotes the scale factor at the end of inflation and $\tilde{\delta}(k, 1)$ is the linearly extrapolated density contrast today. This transfer function has a crucial length-scale, namely $\lambda_\mathrm{eq} = d_\mathrm{H}(a_\mathrm{eq})$, the comoving horizon size at matter-radiation equality. This length-scale can be found in the location of the peak of the matter power spectrum (as seen in Fig. 2.4).

2.1.2.4. Statistics of random fields: correlation functions and polyspectra

In the limit of linear structure formation, we can now describe the evolution of density fluctuations over time. However, investigating the density field itself is not particularly useful if we want to gain information about our Universe. For example, an observer in our Universe that measures the local density contrast at the same cosmic time as we do, just at a different position in space, will see a completely different distribution of density fluctuations. However, the *statistical properties* of these fluctuations remain the same.[10] Throughout this thesis, we assume that we work with a zero-mean field, which holds by definition for the density contrast δ.

The most commonly used statistical property is the two-point correlation function and its Fourier-space analogue, the power spectrum. The two-point correlation function of a field of density perturbations $\delta(\boldsymbol{x})$ is defined as

$$\xi(\boldsymbol{x}, \boldsymbol{y}) = \langle \delta(\boldsymbol{x}) \delta(\boldsymbol{y}) \rangle \ . \tag{2.45}$$

Here, the brackets $\langle \ldots \rangle$ denote an ensemble average. Unfortunately, we only have one Universe to utilise for our statistical analysis, which makes taking an ensemble average difficult. We thus make use of an ergodic-type hypothesis and replace the ensemble average with a volume average (Ellis et al., 2012, page 307). As our Universe is (presumably) homogeneous and isotropic, the value of the correlation function $\xi(\boldsymbol{x}, \boldsymbol{y})$ only depends on the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, meaning $\xi(\boldsymbol{x}, \boldsymbol{y}) = \xi(|\boldsymbol{y} - \boldsymbol{x}|)$.

[10] The same holds for all other observations, like the CMB.

In our cosmological model, it is much easier to describe the Fourier modes $\tilde{\delta}(\boldsymbol{k})$ of the density fluctuations

$$\tilde{\delta}(\boldsymbol{k}) = \int_{\mathbb{R}^3} \mathrm{d}^3x\, \delta(\boldsymbol{x}) \exp(-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}) \tag{2.46}$$

$$\delta(\boldsymbol{x}) = \int_{\mathbb{R}^3} \frac{\mathrm{d}^3k}{(2\pi)^3}\, \tilde{\delta}(\boldsymbol{k}) \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x})\ . \tag{2.47}$$

We can then define the power spectrum as

$$(2\pi)^3\, \delta_\mathrm{D}(\boldsymbol{k}+\boldsymbol{k}')\, P(k) = \left\langle \tilde{\delta}(\boldsymbol{k})\, \tilde{\delta}(\boldsymbol{k}') \right\rangle\ , \tag{2.48}$$

where δ_D is the Dirac delta functional and $k = |\boldsymbol{k}|$ is the modulus of the $\boldsymbol{k}$-vector. The power spectrum P can be expressed as the Fourier transform of the correlation function

$$P(|\boldsymbol{k}|) = \int_{\mathbb{R}^3} \mathrm{d}^3x\, \xi(x) \exp(-\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{k})\ . \tag{2.49}$$

The next-order correlation function is the three-point correlation function:

$$\zeta(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \zeta(|\boldsymbol{z}-\boldsymbol{y}|, |\boldsymbol{z}-\boldsymbol{x}|, |\boldsymbol{y}-\boldsymbol{x}|) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{y})\delta(\boldsymbol{z})\rangle\ . \tag{2.50}$$

Its Fourier-space counterpart is the bispectrum

$$(2\pi)^3 \delta_\mathrm{D}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k_3}) B(k_1,k_2,k_3) = \left\langle \tilde{\delta}(\boldsymbol{k}_1)\tilde{\delta}(\boldsymbol{k}_2)\tilde{\delta}(\boldsymbol{k}_3)\right\rangle\ . \tag{2.51}$$

In general, one can define the n-th order correlations and polyspectra as

$$\begin{aligned} \left\langle \prod_{i=1}^{n} \delta(\boldsymbol{x}_i) \right\rangle &= \xi_n(\boldsymbol{x}_1,\ldots,\boldsymbol{x_n}) \\ \left\langle \prod_{i=1}^{n} \tilde{\delta}(\boldsymbol{k}_i) \right\rangle &= (2\pi)^3\, \delta_\mathrm{D}\left(\sum_{i=1}^{n} \boldsymbol{k}_i\right) P_n(\boldsymbol{k}_1,\ldots,\boldsymbol{k}_n)\ . \end{aligned} \tag{2.52}$$

In the framework of gravitational lensing, we mostly work with two-dimensional fields instead of the three-dimensional ones we are discussing here. All the introduced concepts still apply; one just has to replace the $(2\pi)^3$ with a $(2\pi)^2$ to account for the correct normalisation of the Fourier transform.

To assess the full statistical properties of a random field, knowledge of all n-point correlation functions would be required, which is obviously impossible. However, there is strong evidence that the information content of higher-order correlation functions diminishes in our Universe.

Gaussian and lognormal random fields A random field $\delta(\boldsymbol{x})$ with power spectrum $P(k)$ is called a *Gaussian random field* if its Fourier modes $\tilde{\delta}(\boldsymbol{k})$ are statistically independent and the probability distribution for a single Fourier mode is characterized by a normal distribution with zero mean and variance $P(|\boldsymbol{k}|)$.

For a Gaussian random field, Isserlis' theorem (also called Wick's theorem) states that products of more than two expectation values factorise as

$$\langle X_1 \dots X_n \rangle = \frac{1}{2} \sum_{N \in \mathbb{P}([1,\dots,n])} \left\langle \prod_{i \in N} X_i \right\rangle \left\langle \prod_{j \notin N} X_j \right\rangle \ , \tag{2.53}$$

where $\mathbb{P}([1,\dots,n])$ denotes the *power set* of $[1,\dots,n]$, meaning the set of all its possible sub-sets. As an example, the fourth order correlation between four random variables decomposes as

$$\begin{aligned} \langle X_1 X_2 X_3 X_4 \rangle = \cancelto{0}{\langle X_1 \rangle} \langle X_2 X_3 X_4 \rangle + \cancelto{0}{\langle X_2 \rangle} \langle X_1 X_3 X_4 \rangle + \cancelto{0}{\langle X_3 \rangle} \langle X_1 X_2 X_4 \rangle + \cancelto{0}{\langle X_4 \rangle} \langle X_1 X_2 X_3 \rangle \\ + \langle X_1 X_2 \rangle \langle X_3 X_4 \rangle + \langle X_1 X_3 \rangle \langle X_2 X_4 \rangle + \langle X_1 X_4 \rangle \langle X_2 X_3 \rangle \ . \end{aligned} \tag{2.54}$$

For a zero-mean Gaussian random field, this implies that all odd correlations vanish, $\langle X_1 \dots X_{2n+1} \rangle = 0$, and all even higher-order correlations are already fully defined by the second-order correlations. This implies that the entire information content of a Gaussian random field is captured by its two-point correlation function or its power spectrum.

For an arbitrary random field, we define the *connected correlation* $\langle \dots \rangle_\mathrm{c}$ as the part that remains when the permutations of sub-correlations are subtracted. For example, for a product of four expectation values $\langle X_1 X_2 X_3 X_4 \rangle$, we define its connected part as

$$\langle X_1 X_2 X_3 X_4 \rangle_\mathrm{c} = \langle X_1 X_2 X_3 X_4 \rangle - \langle X_1 X_2 \rangle \langle X_3 X_4 \rangle - \langle X_1 X_3 \rangle \langle X_2 X_4 \rangle - \langle X_1 X_4 \rangle \langle X_2 X_3 \rangle \ . \tag{2.55}$$

In particular, for a Gaussian random field, $\langle \dots \rangle_\mathrm{c} = 0$ for all correlations higher than second-order.

If we want to describe a field that also contains higher-order correlations, we can resort to lognormal fields.[11] Let $\delta_\mathrm{G}(\boldsymbol{x})$ be a Gaussian random field with correlation function $\xi_\mathrm{G}(x)$, zero mean, and unit dispersion (meaning $\xi_\mathrm{G}(0) = 1$). Then the lognormal field $\delta_\mathrm{L}(\boldsymbol{x})$ defined via

$$\delta_\mathrm{L}(\boldsymbol{x}) = A \mathrm{e}^{-\frac{\alpha^2}{2}} \left(\mathrm{e}^{\alpha\, \delta_\mathrm{G}(\boldsymbol{x})} - \mathrm{e}^{\frac{\alpha^2}{2}} \right) \ , \tag{2.56}$$

with

$$A = \frac{\sigma}{\sqrt{\mathrm{e}^{\alpha^2} - 1}} \ , \tag{2.57}$$

has zero mean and dispersion σ. The parameter α describes the degree of non-Gaussianity. Its two-point correlation function can be calculated as (P. Schneider, private comm.)

$$\xi_\mathrm{L}(x) = A^2 \left(\mathrm{e}^{\alpha^2\, \xi_\mathrm{G}(x)} - 1 \right) \ , \tag{2.58}$$

and its three-point correlation function ζ_L can be calculated as (P. Schneider, private comm.)

$$\zeta_\mathrm{L}(x,y,z) = \frac{\xi_\mathrm{L}(x)\, \xi_\mathrm{L}(y)\, \xi_\mathrm{L}(z)}{A^3} + \frac{\xi_\mathrm{L}(x)\, \xi_\mathrm{L}(y) + \xi_\mathrm{L}(x)\, \xi_\mathrm{L}(z) + \xi_\mathrm{L}(y)\, \xi_\mathrm{L}(z)}{A} \ . \tag{2.59}$$

[11] Indeed, it appears that lognormal fields constitute a reasonably good approximation for the late-time density field of the Universe (Hilbert et al., 2011).

2.1.2.5. The power spectrum in linear theory

An overwhelming amount of evidence suggests that the CMB is at least very close to a Gaussian random field.[12] It is essential to note that, within the framework of linear structure formation, the spatial and temporal dependencies of the density field decouple (see Eq. 2.41), meaning that the amplitudes of density perturbations grow, but the structure of the field itself does not change. In particular, this means that if the initial conditions are a Gaussian random field, then at all times during linear structure formation, the density field remains Gaussian.

This implies that the density perturbations of the early Universe have a vanishing bispectrum and that the knowledge of their power spectrum already provides us with all the information we can get about our cosmological model. Models of inflation predict that the initial post-inflation power spectrum is a power law $P_\mathrm{i}(k) \propto k^{n_s}$ with spectral index $n_s \lesssim 1$. From our previous considerations, we can thus define the linear power spectrum as

$$P_\mathrm{lin}(k, a) = A\, T_k^2\, D_+^2(a)\, k^{n_s} \ . \tag{2.60}$$

The normalisation constant A can also be defined by the parameter σ_8, which measures the variance of linear matter fluctuations within a comoving radius of $8\,h^{-1}\,\mathrm{Mpc}$. It can be calculated from the quantity

$$\sigma^2(R) = \frac{1}{2\pi^2} \int_0^\infty \mathrm{d}k\, k^2\, P_\mathrm{lin}(k, a=1)\, \tilde{W}^2(kR) \ , \tag{2.61}$$

where $\tilde{W}(x) = 3\,(\sin x - x\,\cos x)/x^2$ is the Fourier transform of a top-hat filter, by setting $R = 8\,h^{-1}\,\mathrm{Mpc}$. The scale of $8\,h^{-1}\,\mathrm{Mpc}$ was chosen such that σ_8 is approximately unity. We see that $\sigma_8^2 \propto A$ holds, so fixing either of them uniquely defines the amplitude of the power spectrum.

We note that this picture is a bit simplified since it neglects the effects of baryons, which were coupled to radiation in the pre-CMB-era. For an accurate model of the linear power spectrum, one would have to use a Boltzmann-solver like Code for Anisotropies in the Microwave Background (CAMB, Lewis and Challinor, 2011). We show the linear matter power spectrum in Fig. 2.4. One can see that the constraints from different observations, spanning a broad range of cosmic scales and times, all agree relatively well with each other, which is a testament to why the ΛCDM model has become the standard model of cosmology.

2.1.2.6. Non-linear structure formation

In the last sections, we have discussed the linear structure formation in our Universe. We know that this formalism can no longer describe the current density field. The existence of stars and planets, which have a density contrast of $\delta \sim 10^{30}$, proves that the assumption $\delta \ll 1$ is not justified everywhere. Even on large scales, galaxy clusters have an average density contrast that surpasses unity by about two orders of magnitude. We thus know that our linear approximation breaks down at some point, and we need to find a formalism to describe the non-linear regime of structure formation.

[12] Searches for non-Gaussianities in the CMB, which could constrain different scenarios of inflation, are still ongoing.

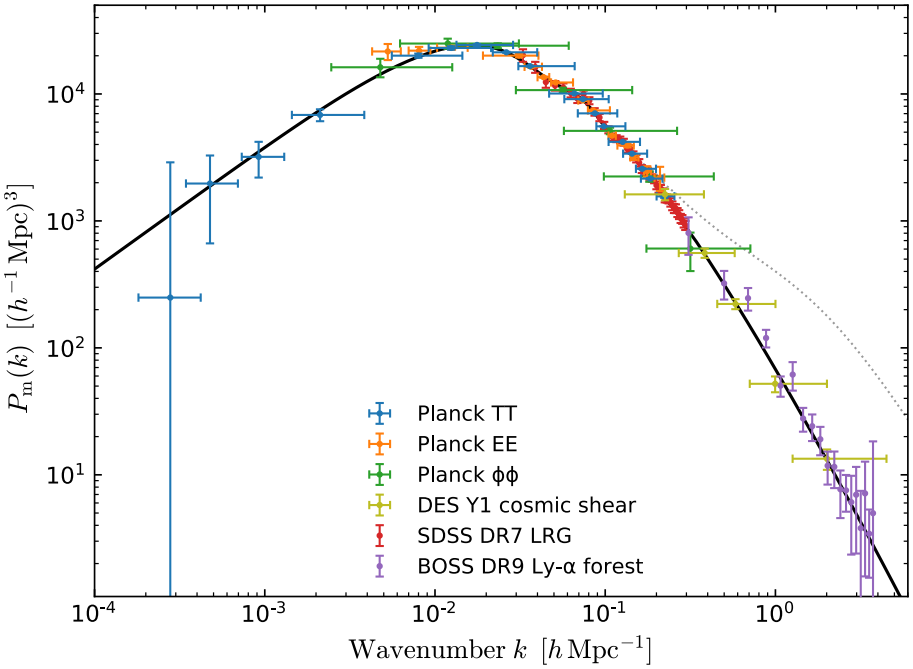


Figure 2.4. – The linear matter power spectrum at $z = 0$ determined by different probes. The dotted black line shows the impact of non-linear structure formation, which we discuss in Sect. 2.1.2.6. From Planck Collaboration et al. (2020a).

Our arguably best chance to accurately model this is by cosmological N-body simulations. These simulations are conducted in a three-dimensional cube of comoving side-length L. First, a Gaussian random field is created from a linear power spectrum at a time where the density distribution of the Universe can still be considered Gaussian (usually $a \sim 0.01$)[13]. Then, discrete particles are distributed in this box according to the probability density dictated by the random field. These particles usually have a mass of $M \sim 10^9 M_\odot$, which is crucial for the mass-resolution of the simulation.

After the initial density field is constructed, the gravitational interactions between the particles are modelled in the cubic, comoving box. Accounting for the interactions of all particle pairs is of computational complexity $\mathcal{O}(N^2)$, which would make the computation prohibitively expensive. There are two potential strategies to mitigate this. The first one is a particle-mesh method, in which the density field of the Universe is computed on a regular grid from the individual particles. Then, the gravitational interaction can be computed from a convolution, which can be performed in Fourier space utilising the convolution theorem. This reduces the computational complexity to $\mathcal{O}(N_\mathrm{pix} \log N_\mathrm{pix})$, where N_pix is the number of pixels in the grid, but it introduces inaccuracies on length scales approaching the pixel size. To account for these inaccuracies, one corrects for the small-scale interactions either by pre-computed correction formulae or by computing these interactions directly.

The second method is a tree code. For this, a hierarchical tree is computed from the particles. Each cell of the tree contains a set of galaxies and represents their total mass

[13]Some simulations also use Lagrangian perturbation theory to set the initial conditions.

and (weighted) average position. Each cell is then recursively sub-divided into two subcells until a pre-defined maximum sub-division is achieved. To model large-scale gravitational interactions between individual particles, one can now compute the interactions between the respective cells containing these particles, which yields an excellent approximation. The closer two particles are together, the smaller the cells used for this computation have to be.

Usually, N-body simulations only contain dark matter since it is the dominant matter component of the Universe and is relatively easy to model, as it only interacts gravitationally. Simulations also containing baryons are much more complex, as they also need to model all kinds of baryonic effects, from radiative cooling of gas to complex feedback processes of supernovae or supermassive black holes. Neglecting these baryons comes at a cost. Firstly, we might know the density distribution of dark matter after running an N-body simulation, but we do not know at which points in space we would observe galaxies or what the properties of these galaxies would be. In this work, we analyse the signal from cosmic shear, which does not distinguish between baryonic and dark matter, so this is not an issue for us. However, due to gravitational interaction, the baryonic matter distribution also affects the dark matter distribution. The abovementioned feedback processes modify the power spectrum of density fluctuations, especially on small scales. Usually, one tries to mitigate these effects by correction factors that are derived from hydrodynamical simulations.

However, even for gravitational self-interaction, the discretisation of the matter distribution into particles of mass $M \sim 10^9 M_\odot$ leads to close encounters. These result in unphysical scattering processes, which would not happen in a smooth matter distribution. To mitigate this, a force-softening is introduced, where the strength of gravitational interactions is reduced below a specific length scale.

N-body simulations can not be generally trusted for small halos containing less than ~ 100 particles, as the finite mass resolution skews the results. For large halos, the force-softening length has an impact and can influence the density profile. Furthermore, scales larger than $\sim 1/3$ of the box-size L are affected by the periodic boundary conditions and should also not be trusted.

Despite all these limitations, N-body simulations are the most accurate way to model and understand non-linear structure formation. We have no way to verify that the predictions made by these simulations are accurate, but we know that different N-body codes produce very consistent results (Springel et al., 2021), which is a good indicator that we understand the limitations of the simulations well. Of course, it is unfeasible to run N-body simulations whenever we want to predict a statistic, which is why other approaches to non-linear structure formation have been developed.

The most straightforward way to achieve that is to go beyond linear perturbation theory. Lagrangian perturbation theory allows to compute the 1-loop and 2-loop corrections for non-linear effects on the power spectrum. Employing these, it is possible to calculate the power spectrum up to the mildly non-linear regime of $\delta \approx 1$. The behaviour at smaller scales, where the density contrast is much larger than unity, can not be captured by perturbation theory.

Applying leading order perturbation theory yields the tree-level bispectrum

$$\begin{aligned} B_{\mathrm{tree}}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3, a) = &\, 2\, F(\boldsymbol{k}_1, \boldsymbol{k}_2)\, P(k_1, a)\, P(k_2, a) \\ &+ 2\, F(\boldsymbol{k}_1, \boldsymbol{k}_3)\, P(k_1, a)\, P(k_3, a) \\ &+ 2\, F(\boldsymbol{k}_2, \boldsymbol{k}_3)\, P(k_2, a)\, P(k_3, a)\ , \end{aligned} \tag{2.62}$$

with the perturbation kernel

$$F(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{5}{7} + \frac{2}{7}\frac{(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)^2}{k_1^2\, k_2^2} + \frac{1}{2}\frac{\boldsymbol{k}_1 \cdot \boldsymbol{k}_2}{k_1\, k_2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right)\ . \tag{2.63}$$

A more phenomenologically motivated approach is the *halo model* formalism, which assumes that all matter is contained in dark matter halos of an identical density profile, which have collapsed and virialised. This assumption is certainly not true, as a large part of the inter-galactic medium has not collapsed into any dark matter halos. However, the halo-model formalism based on this assumption yields remarkably accurate results for the statistical matter distribution in the non-linear regime.

Given the abovementioned assumption, we need to fix three quantities to get a well-defined model: The density profile of the dark matter halos $\rho(r)$, the number density of halos of mass M at scale factor a, $n(M, a)$, and the spatial distribution of dark matter halos in the Universe. Navarro et al. (1996) found from N-body simulations that dark matter halos all have a similar density profile, the Navarro-Frenk-White (NFW) profile, which is usually used as the density profile in the halo model. The halo mass function was analytically derived by Press and Schechter (1974) under the assumption that dark matter haloes undergo a spherical collapse. Unfortunately, an ellipsoidal collapse of dark matter haloes is a much more likely scenario, so nowadays, a fitting formula from N-body simulations (such as Sheth and Tormen, 1999) is used. The spatial distribution of dark matter halos is subject to a halo bias that describes when overdensities start collapsing into dark matter halos. For our purposes, we require both the linear and quadratic halo bias $b_1(M, a)$ and $b_2(M, a)$

$$\delta_{\mathrm{H}}(\boldsymbol{x}, a|M) \approx b_1(M, a)\, \delta(\boldsymbol{x}, a) + \frac{1}{2} b_2(M, a)\, \delta^2(\boldsymbol{x}, a)\ , \tag{2.64}$$

which can be approximated by equations 14 and 15 of Scoccimarro et al. (2001)

We now briefly describe how Takahashi et al. (2020) construct a bispectrum model from the halo model. The halo model power spectrum can be constructed in the same way. The total matter bispectrum is comprised of a sum of three terms, the 1-halo term, the 2-halo term and the 3-halo term. The 1-halo term describes the contribution of triangle configurations where all three particles reside in the same dark matter halo:

$$B_{1\mathrm{h}}(k_1, k_2, k_3, a) = \int \mathrm{d}M\, \frac{\mathrm{d}n(M, a)}{\mathrm{d}M} \left(\frac{M}{\bar{\rho}(a)}\right)^3 \tilde{u}(k_1, M, a)\, \tilde{u}(k_2, M, a)\, \tilde{u}(k_3, M, a)\ , \tag{2.65}$$

where $\tilde{u}(k, M, a)$ describes the Fourier transform of the scaled density profile $\rho(r, M, a)/M$. The 2-halo term describes the contribution of triangles where two points are in one halo, and the third point is in the second one. It can be computed as

$$B_{2\mathrm{h}}(k_1, k_2, k_3, a) = I_2^1(k_1, k_2, a)\, I_1^1(k_3, a)\, P_{\mathrm{lin}}(k_3, a) + 2 \text{ Perm.}\ , \tag{2.66}$$

with

$$I_1^\alpha(k,a) = \int \mathrm{d}M \, \frac{\mathrm{d}n(M,a)}{\mathrm{d}M} \frac{M}{\bar{\rho}(a)} \, b_\alpha(M,a) \, \tilde{u}(k,M,a)$$
$$I_2^\alpha(k_1,k_2,a) = \int \mathrm{d}M \, \frac{\mathrm{d}n(M,a)}{\mathrm{d}M} \left(\frac{M}{\bar{\rho}(a)}\right)^2 b_\alpha(M,a) \, \tilde{u}(k_1,M,a) \, \tilde{u}(k_2,M,a) \, . \tag{2.67}$$

The 3-halo term can then be computed as

$$B_{3\mathrm{h}}(k_1,k_2,k_3,a) = I_1^1(k_1,a) \, I_1^1(k_2,a) \, I_1^1(k_3,a) \, B_{\mathrm{tree}}(k_1,k_2,k_3,a) + \Big[I_1^1(k_1,a) \, I_1^1(k_2,a) \, I_1^2(k_3,a) \, P_{\mathrm{lin}}(k_1,a) \, P_{\mathrm{lin}}(k_2,a) + 2 \text{ Perm.} \Big] \, . \tag{2.68}$$

Naturally, the 1-halo term dominates for small scales, whereas the 3-halo term dominates for large scales. The 2-halo term is only relevant for squeezed triangles with $k_1 \approx k_2 \gg k_3$ and sub-dominant on all scales otherwise. While the BiHalofit algorithm (Takahashi et al., 2020) utilises the functional form of the halo model, it replaces the $\tilde{u}$ and the I_β^α from Eq. (2.67) by fitting functions. Furthermore, it removes the 2-halo term and accounts for that by replacing the linear power spectrum in the 3-halo term with an 'enhanced' linear power spectrum, which is also determined by fitting parameters. This method appears to be more accurate than finding appropriate functions for the density profile, number density, and spatial distribution of dark matter halos separately.

Besides giving a functional form for more accurate fitting formulae, the halo model's advantage is that higher-order polyspectra can also be computed from this formalism with relative ease. For example the 1-halo term of any n-th order polyspectrum (see Eq. 2.52), can be computed as

$$P_{n,1\mathrm{h}}(k_1,\ldots,k_n,a) = \int \mathrm{d}M \, \frac{\mathrm{d}n(M,a)}{\mathrm{d}M} \left(\frac{M}{\bar{\rho}(a)}\right)^n \prod_{i=1}^{n} \tilde{u}(k_i,M,a) \, . \tag{2.69}$$

2.2. Gravitational lensing

Gravitational lensing is a powerful tool that can be used in various cosmological applications. Using gravitational lensing, we can determine the masses of galaxy clusters (see e.g. Applegate et al., 2016; Schrabback et al., 2018, 2021), the value of the Hubble constant (e.g. Wong et al., 2020), analyse the abundance of compact 'dark' objects in the Universe (e.g. Alcock et al., 2000) and determine the cosmological parameter $S_8 = \sigma_8 \sqrt{\Omega_\mathrm{m}/0.3}$, which is defined along the main degeneracy direction for weak lensing studies. For the latter, we have discovered a tension between the local Universe and the CMB (Hildebrandt et al., 2017; Planck Collaboration et al., 2020b; Joudaki et al., 2020; Heymans et al., 2021; Abbott et al., 2022; Di Valentino et al., 2021b), which might give insights into as-of-yet undiscovered physical processes (Abdalla et al., 2022). This tension is not (yet) as pronounced as the Hubble constant tension: Planck Collaboration et al. (2020b) report $S_8 = 0.834 \pm 0.016$ from CMB observations; from weak gravitational lensing we get a 3σ tension in KiDS ($S_8 = 0.759^{+0.024}_{-0.021}$ in Asgari et al., 2021) and a $2.1\text{-}2.3\,\sigma$ tension in DES ($S_8 = 0.759^{+0.025}_{-0.023}$ or $S_8 = 0.772^{+0.018}_{-0.017}$ in Secco et al., 2022b; Amon et al., 2022). Different

surveys and summary statistics all yield values of S_8 that are about $1.5\text{-}3\,\sigma$ lower than the one determined from the CMB.

One of the main advantages of the gravitational lensing effect is that it is sensitive to both baryonic and dark matter; to quantify this effect in models, we thus do not have to understand the complicated processes of galaxy formation and evolution. In this section, we briefly review the formalism that we use throughout the rest of this thesis. This section is heavily based on Bartelmann and Schneider (2001); Schneider (2005); Schneider (2012).

We have already discussed in Sect. 2.1.1.1 that light-rays follow geodesics, or straight lines, through the 4-dimensional spacetime manifold. However, the metric of that manifold is distorted by the presence of mass, leading to a curved spacetime. Therefore, even though light describes geodesics in the 4-dimensional spacetime, the light-rays projected to the three spatial dimensions are curved by the presence of mass.

Assuming a small gravitational potential $|\Phi|/c^2 \ll 1$, the deflection of a light-ray by a point-mass M at impact paramter ξ can be calculated by linearising the field equations (2.1) and yields

$$\hat{\alpha} = \frac{4GM}{c^2\xi} \,. \tag{2.70}$$

Curiously, when one computes the deflection of a particle of mass $m \ll M$ passing by with velocity v at the same impact parameter ξ in Newtonian gravity, one receives

$$\hat{\alpha} = \frac{2GM}{v^2\xi} \,. \tag{2.71}$$

One can see that this deflection angle does not depend on the mass m, so naively setting $m = 0$ and $v = c$ for photons, we receive a deflection angle half as large as the one predicted by GR. An observation of the deflection of observed stellar positions by the Sun during a solar eclipse (Dyson et al., 1920) is widely regarded as the first significant confirmation of the predictions of GR.

2.2.1. The gravitational lensing formalism

In principle, gravitational lensing is only correctly described in the framework of GR, but a few simplifying assumptions ease the complexity to a manageable level. As mentioned above, the first assumption is that the gravitational potential Φ is small, meaning $|\Phi|/c^2 \ll 1$. This assumption is highly justified; even in the central regions of galaxy clusters, the gravitational potential is only about $|\Phi| \sim 10^{-5}\,c^2$. The only place this assumption breaks down is in the vicinity of neutron stars and black holes. This enables us to make two critical simplifications. Firstly, we can describe the deflection angle by a point mass via Eq. (2.70). Secondly, and even more importantly, we can treat deflections by different masses as independent. This means that for an ensemble of masses, the combined deflection angle is the sum of the individual deflection angles. Without these assumptions, the computational complexity for even the simple case of a point-mass lens is enormous (James et al., 2015).

We further assume that the velocities of the observer, lens and source are small, $v \ll c$, and that the involved length scales L are small compared to the Hubble radius, $L \ll c/H_0$. Among other things, this means that we can neglect effects that arise because the observer and source reside in different rest-frames. A gauge-invariant lensing formalism

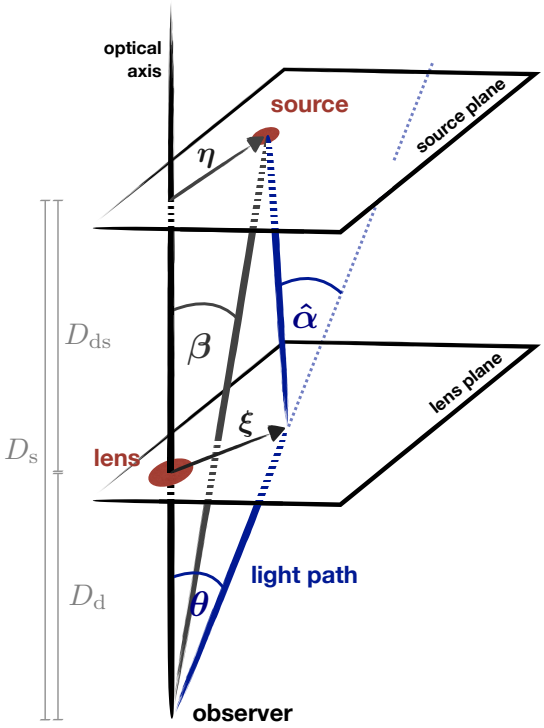


Figure 2.5 – A sketch of a gravitational lensing system. Figure adapted from Bartelmann and Schneider (2001) by S. Unruh.

has been developed (Yoo et al., 2018; Grimm and Yoo, 2018), but the standard weak lensing approach appears to be sufficient except for very large scales (Giblin et al., 2017; Grimm and Yoo, 2021).

We further assume that all angles involved in the calculations are small, such that $\sin(\alpha) \approx \tan(\alpha) \approx \alpha$ holds, which it does to an accuracy of 10^{-3} at 5° and 10^{-4} at 1°. For now, we also want to make the *thin lens approximation*, meaning that all matter responsible for the lensing effect is assumed to be at the same distance from the observer.

Under these assumptions, we can sketch an example lensing system in Fig. 2.5. Depicted is a source at position $\boldsymbol{\eta} = D_{\mathrm{s}}\boldsymbol{\beta}$, which emits light that gets deflected by the angle $\hat{\boldsymbol{\alpha}}$ and is thus observed to be at position $\boldsymbol{\xi} = D_{\mathrm{d}}\boldsymbol{\theta}$. Note that the distances $D_{\mathrm{d}}, D_{\mathrm{ds}}$ and D_{s} are angular diameter distances and thus, contrary to what the sketch might suggest, $D_{\mathrm{d}} + D_{\mathrm{ds}} \neq D_{\mathrm{s}}$ holds. We can use Thales's theorem to derive the *lens equation*

$$\boldsymbol{\beta}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\hat{\boldsymbol{\alpha}}(D_{\mathrm{d}}\theta) = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}) \,, \tag{2.72}$$

where we have defined the scaled deflection angle

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\,\hat{\boldsymbol{\alpha}}(D_{\mathrm{d}}\boldsymbol{\theta}) \,. \tag{2.73}$$

For a mass distribution ρ, we can define the surface mass density

$$\Sigma(\boldsymbol{\xi}) = \int \mathrm{d}r_3\, \rho(\xi_1, \xi_2, r_3) \,, \tag{2.74}$$

as the 2-dimensional projection of the density along the line of sight. The deflection angle $\boldsymbol{\alpha}$ can then be computed as

$$\begin{aligned}\boldsymbol{\alpha}(\boldsymbol{\theta}) &= \frac{4G}{c^2}\frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\int \mathrm{d}^2\xi'\,\Sigma(\boldsymbol{\xi}')\,\frac{\boldsymbol{\xi}-\boldsymbol{\xi}'}{|\boldsymbol{\xi}-\boldsymbol{\xi}'|^2} \\ &= \frac{1}{\pi}\int \mathrm{d}^2\theta'\,\kappa(\boldsymbol{\theta}')\,\frac{\boldsymbol{\theta}-\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}\ ,\end{aligned} \tag{2.75}$$

where we have defined the dimensionless *convergence* κ as

$$\kappa(\boldsymbol{\theta}) = \frac{4\pi G}{c^2}\frac{D_{\mathrm{d}}D_{\mathrm{ds}}}{D_{\mathrm{s}}}\Sigma(D_{\mathrm{d}}\boldsymbol{\theta}) = \frac{\Sigma(D_{\mathrm{d}}\boldsymbol{\theta})}{\Sigma_{\mathrm{crit}}}\ , \tag{2.76}$$

with the *critical surface mass density*

$$\Sigma_{\mathrm{crit}} = \frac{c^2}{4\pi G}\frac{D_{\mathrm{s}}}{D_{\mathrm{d}}D_{\mathrm{ds}}}\ . \tag{2.77}$$

When we define the lensing potential

$$\Psi(\boldsymbol{\theta}) = \frac{1}{\pi}\int \mathrm{d}\theta'\,\kappa(\boldsymbol{\theta}')\,\ln\left(|\boldsymbol{\theta}-\boldsymbol{\theta}'|\right)\ , \tag{2.78}$$

then we can see that both deflection angle and convergence can be written as derivatives of that potential, namely

$$\boldsymbol{\alpha} = \nabla\Psi, \quad \kappa = \frac{\nabla^2\Psi}{2}\ . \tag{2.79}$$

We then compute the Jacobian of the lens equation, $\mathcal{A}_{ij} = \partial\beta_i/\partial\theta_j$, as

$$\mathcal{A}(\boldsymbol{\theta}) = \left(\delta_{ij} - \frac{\partial^2\Psi(\boldsymbol{\theta})}{\partial\theta_i\,\partial\theta_j}\right) = \begin{pmatrix}1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1\end{pmatrix} = (1-\kappa)\begin{pmatrix}1-g_1 & -g_2 \\ -g_2 & 1+g_1\end{pmatrix}\ , \tag{2.80}$$

where we define the complex *shear* in cartesian coordinates γ_{c} as

$$\gamma_{\mathrm{c}} = \gamma_1 + \mathrm{i}\gamma_2 = \frac{1}{2}(\Psi_{,11} - \Psi_{,22}) + \mathrm{i}\Psi_{,12}\ , \tag{2.81}$$

and the *reduced shear* g as $g_i = \gamma_i/(1-\kappa)$.

Assuming that a source galaxy at observed position $\boldsymbol{\theta}_0$ is small, we can set $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$ and linearise the lens equation as

$$\boldsymbol{\beta}(\boldsymbol{\theta}) - \boldsymbol{\beta}_0 = \mathcal{A}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\ . \tag{2.82}$$

From inspection of the Jacobian $\mathcal{A}$, we see that the convergence κ leads to an isotropic scaling of a source with the factor $(1-\kappa)$. On the other hand, the terms g_1 and g_2 lead to an anisotropic distortion of the image.

Furthermore, we can look at an image's magnification μ. Since photons are not created or destroyed in gravitational lensing, the magnification is simply the ratio of the sizes between the lensed and un-lensed image:

$$\mu = \frac{\mathrm{d}^2\beta}{\mathrm{d}^2\theta} = \frac{1}{|\det(\mathcal{A})|} = \frac{1}{(1-\kappa)^2 - |\gamma_{\mathrm{c}}|^2}\ . \tag{2.83}$$

We see that there is a possibility that the magnification diverges because the matrix $\mathcal{A}$ becomes singular. In this case, the lens equation is no longer a one-to-one mapping; the same source can appear at multiple positions in the sky. This usually happens when $\kappa \gtrsim 1$ holds. The convergence thus presents an ideal measure to distinguish *weak gravitational lensing* with $\kappa \ll 1$ from *strong gravitational lensing* with $\kappa \gtrsim 1$. In weak gravitational lensing, we assume that the Jacobian $\mathcal{A}$ does not become singular[14]. The inverse function theorem then implies that in weak gravitational lensing, multiple images of the same source can not occur since the lens equation is a one-to-one mapping.

Our main observable in weak gravitational lensing is the ellipticity ϵ. Assuming a source galaxy with perfectly elliptical isophotes, we can define its ellipticity as

$$\epsilon = \frac{a-b}{a+b} \exp(2\mathrm{i}\varphi) \ , \tag{2.84}$$

where a and b are the semi-major- and minor-axes of the ellipse and φ is the angle between a galaxy's semi-major axis a and the x-axis.[15] An observed source galaxy has some intrinsic ellipticity ϵ_{int} in addition to the reduced shear g it experiences due to the weak lensing effect. The observed ellipticity can then be calculated as

$$\epsilon_{\mathrm{obs}} = \frac{\epsilon_{\mathrm{int}} + g}{1 + g^* \epsilon_{\mathrm{int}}} \ . \tag{2.85}$$

The intrinsic ellipticity of a galaxy is usually much greater than the shear signal. However, the cosmological principle states that there is no preferred direction in the Universe, meaning that for the average intrinsic ellipticity $\langle \epsilon_{\mathrm{int}} \rangle = 0$ holds. If we further assume that the intrinsic ellipticity is not correlated with the shear signal, the observed ellipticity becomes an unbiased estimator of the reduced shear

$$\langle \epsilon_{\mathrm{obs}} \rangle = \langle \epsilon_{\mathrm{int}} \rangle + \langle g \rangle = \langle g \rangle \ . \tag{2.86}$$

In weak lensing, we usually assume that $\gamma \approx g$ holds, as κ is small.

Even though the convergence κ is not directly observable, it can, in principle, be reconstructed from a shear field. The most straight-forward way uses the fact that the Fourier transforms of the convergence and shear are related (Kaiser and Squires, 1993)

$$\tilde{\gamma}_{\mathrm{c}}(\boldsymbol{\ell}) = \mathrm{e}^{2\mathrm{i}\varphi_\ell} \tilde{\kappa}(\boldsymbol{\ell}) \ , \tag{2.87}$$

where φ_ℓ is the polar angle of the $\boldsymbol{\ell}$-vector. In real space, we can write this equation as a convolution:

$$\begin{aligned} \gamma_{\mathrm{c}}(\boldsymbol{\theta}) &= \frac{1}{\pi} \int_{\mathbb{R}^2} \mathrm{d}^2\theta' \, \mathcal{D}(\boldsymbol{\theta} - \boldsymbol{\theta'}) \, \kappa(\boldsymbol{\theta'}) \ , \\ \kappa(\boldsymbol{\theta}) &= \kappa_0 + \frac{1}{\pi} \int_{\mathbb{R}^2} \mathrm{d}^2\theta' \, \mathcal{D}^*(\boldsymbol{\theta} - \boldsymbol{\theta'}) \, \gamma_{\mathrm{c}}(\boldsymbol{\theta'}) \ , \text{with} \\ \mathcal{D}(\boldsymbol{\theta}) &= \ -(\theta_1 - \mathrm{i}\theta_2)^{-2} \ . \end{aligned} \tag{2.88}$$

[14] Of course, in each weak gravitational lensing survey, there are points where $\kappa > 1$ holds, for example in the central region of a massive galaxy cluster. However, compared to the total survey area, these points are sporadic.

[15] Naturally, the images of real galaxies are never perfectly elliptical and assigning an ellipticity to these images is more difficult. There are several different estimators for the ellipticity of a source galaxy, from moments-based methods to template fitting or machine-learning-based methods.

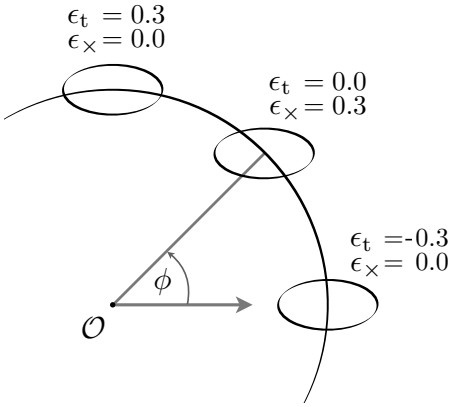


Figure 2.6 – An example illustration of tangential and cross shear for a galaxy with $\gamma_1 = 0.3$, $\gamma_2 = 0$ in cartesian coordinates. Figure adapted from work by M. Bradac in Schneider (2012) by S. Unruh.

Here, κ_0 is an arbitrary constant. The disadvantage of this method is that replacing the integral with a sum over an ensemble of observed galaxy ellipticities yields an estimator with infinite noise, so some smoothing needs to be applied. Furthermore, this method suffers from a bias in the presence of masks or gaps in the data. Better mass-reconstruction methods have been developed (Seitz and Schneider, 2001; Hong et al., 2021) that mitigate these issues. However, one critical flaw is the so-called *mass-sheet degeneracy* (Falco et al., 1985; Gorenstein et al., 1988; Schneider and Seitz, 1995; Schneider and Sluse, 2013), which states that no lensing observable is modified by the transformation[16]

$$\kappa_\lambda(\boldsymbol{\theta}) = \lambda\kappa(\boldsymbol{\theta}) + (1-\lambda), \quad \beta_\lambda = \lambda\beta \; . \tag{2.89}$$

In this thesis, we work with the *aperture mass statistics* $\mathcal{M}_{\rm ap}$, which can be seen as a smoothed and compensated convergence field. To introduce these statistics, we define the *tangential* and *cross* shear $\gamma_{\rm t,\times}(\boldsymbol{\theta};\boldsymbol{\theta}')$ with respect to an orientation $\boldsymbol{\theta}'-\boldsymbol{\theta}$ as (see Fig. 2.6)

$$\gamma_{\rm t}(\boldsymbol{\theta};\boldsymbol{\theta}'-\boldsymbol{\theta}) + {\rm i}\gamma_\times(\boldsymbol{\theta};\boldsymbol{\theta}'-\boldsymbol{\theta}) = -\gamma_{\rm c}(\boldsymbol{\theta})\exp(-2{\rm i}\varphi) = -\gamma_{\rm c}(\boldsymbol{\theta})\frac{(\boldsymbol{\theta}'-\boldsymbol{\theta})^*}{\boldsymbol{\theta}'-\boldsymbol{\theta}} \; , \tag{2.90}$$

where φ is the polar angle of the vector $\boldsymbol{\theta}'-\boldsymbol{\theta}$. It can be shown that overdensities in the matter distribution induce a positive tangential shear around them and that the average cross shear on a circle around any point is zero. In light of this, we can introduce the aperture mass. Let $U_{\theta_{\rm ap}}(\theta)$ be a compensated filter with $\int {\rm d}\theta\, \theta\, U_{\theta_{\rm ap}}(\theta) = 0$. Here, $\theta_{\rm ap}$ is a scaling parameter of the $U_{\theta_{\rm ap}}$-filter and is called the *aperture radius*. The aperture mass $\mathcal{M}_{\rm ap}(\boldsymbol{\theta};\theta_{\rm ap})$ at point $\boldsymbol{\theta}$ is then defined as

$$\mathcal{M}_{\rm ap}(\boldsymbol{\theta};\theta_{\rm ap}) = \int {\rm d}^2\theta'\, \kappa(\boldsymbol{\theta}')\, U_{\theta_{\rm ap}}(|\boldsymbol{\theta}-\boldsymbol{\theta}'|) \; . \tag{2.91}$$

The fact that the filter $U_{\theta_{\rm ap}}$ is compensated has a major advantage: a uniform convergence sheet κ_0 does not affect the aperture mass statistic. Furthermore, the aperture statistics are also a zero-mean field, $\langle\mathcal{M}_{\rm ap}(\boldsymbol{\theta};\theta_{\rm ap})\rangle_{\boldsymbol{\theta}} = 0$. The aperture mass statistic can also be constructed from a shear field via

$$\mathcal{M}_{\rm ap}(\boldsymbol{\theta};\theta_{\rm ap}) + {\rm i}\mathcal{M}_\perp(\boldsymbol{\theta};\theta_{\rm ap}) = \int {\rm d}^2\theta'\, Q_{\theta_{\rm ap}}(|\boldsymbol{\theta}-\boldsymbol{\theta}'|)\left[\gamma_{\rm t}(\boldsymbol{\theta};\boldsymbol{\theta}'-\boldsymbol{\theta}) + {\rm i}\gamma_\times(\boldsymbol{\theta};\boldsymbol{\theta}'-\boldsymbol{\theta})\right] \; , \tag{2.92}$$

[16] To be precise, the image positions, magnification ratios and image shapes remain unaffected, whereas the time-delay Δt between two images of the same source is modified by a factor of λ.

with the filter function

$$Q_{\theta_{\rm ap}}(\theta) = \frac{2}{\theta^2}\int_0^\theta \mathrm{d}\theta'\,\theta' U_{\theta_{\rm ap}}(\theta') - U_{\theta_{\rm ap}}(\theta)\;. \tag{2.93}$$

Throughout this work, we assume the following filter functions (Crittenden et al., 2002):

$$Q_{\theta_{\rm ap}}(\theta) = \frac{\theta^2}{4\pi{\theta_{\rm ap}}^4}\exp\left(-\frac{\theta^2}{2{\theta_{\rm ap}}^2}\right);\quad U_{\theta_{\rm ap}}(\theta) = \frac{1}{2\pi{\theta_{\rm ap}}^2}\left(1-\frac{\theta^2}{2{\theta_{\rm ap}}^2}\right)\exp\left(-\frac{\theta^2}{2{\theta_{\rm ap}}^2}\right)\;. \tag{2.94}$$

Apart from an estimator with finite noise properties, aperture mass statistics have another undeniable advantage: The fact that the average cross shear on a circle around any point in the Universe is zero means that a significant non-zero $\mathcal{M}_\perp$ is a smoking gun for something going wrong in the measurements.

2.2.2. E- and B-modes in gravitational lensing

The careful reader may have noticed something odd in the convergence reconstruction, particularly in Eq. (2.88). The convergence κ is, by definition, a real quantity, whereas the convergence reconstruction is achieved by convolving the complex shear with a complex filter function. In the optimal case, meaning a valid lensing formalism and a noise-free shear field, the imaginary part of the reconstructed convergence is indeed zero. In practice, an imaginary part of the convergence emerges in the presence of shape noise. We thus separate the convergence into *E- and B-modes* (Schneider et al., 2002)

$$\kappa = \kappa_{\rm E} + \mathrm{i}\kappa_{\rm B}\;. \tag{2.95}$$

Here, $\kappa_{\rm B}$ can, to first order, not be created by any gravitational lensing effect. A non-negligible measurement of $\kappa_{\rm B}$ is thus a sign that either one of the used assumptions, such as the reduced shear approximation, breaks down or that an untreated systematic effect is modifying the shear measurements. Expanding Eq. (2.91) to

$$\mathcal{M}_{\rm ap}(\boldsymbol{\theta};\theta_{\rm ap}) + \mathrm{i}\mathcal{M}_\perp(\boldsymbol{\theta};\theta_{\rm ap}) = \int \mathrm{d}^2\theta'\;\left[\kappa_{\rm E}(\boldsymbol{\theta}') + \mathrm{i}\kappa_{\rm B}(\boldsymbol{\theta}')\right] U_{\theta_{\rm ap}}(|\boldsymbol{\theta}-\boldsymbol{\theta}'|)\;, \tag{2.96}$$

shows that $\mathcal{M}_\perp$ is an alternative way to measure the B-modes in gravitational lensing. Looking at the definition of tangential and cross shear (Eq. 2.90) and how they relate to the aperture mass statistic (Eq. 2.92), we can see that tangential shear patterns create E-modes, whereas cross shear patterns create B-modes. A sketch of this is shown in Fig. 2.7.[17]

These B-modes provide a powerful null-test for a gravitational lensing analysis. When any lensing analysis is performed, and the detected B-modes are statistically significant, this points to an error in the analysis pipeline (such as e.g. Hildebrandt et al., 2017). One can then investigate which parts of the analysis can cause B-modes by simulating, for example, leakage of the telescope's point-spread function (PSF) and investigating the B-mode pattern caused by this effect (compare Asgari et al., 2019; Asgari and Heymans, 2019). It should be noted that, while the non-detection of B-modes is an excellent sign, it does not necessarily mean that there are no errors in the analysis pipeline.

[17] It should be noted that there are non-zero shear patterns that are neither E- nor B-modes, for example, a linear shear field. These *ambiguous modes* are discussed in detail in Schneider et al. (2022).

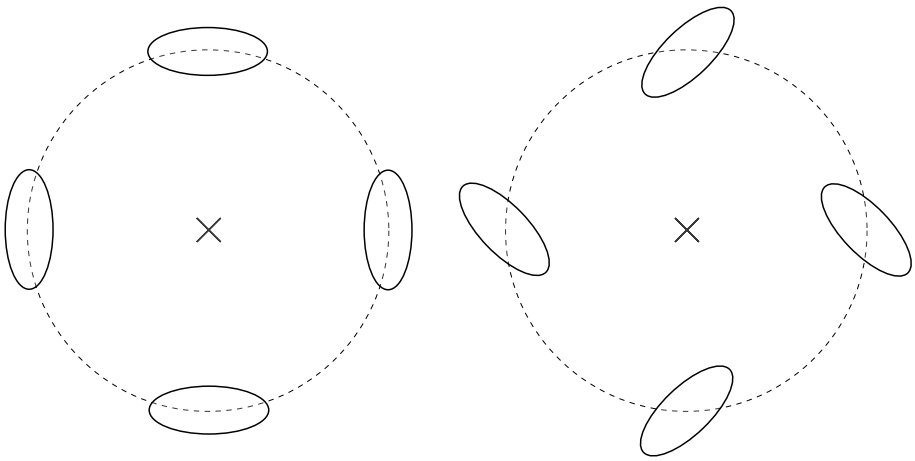


Figure 2.7. – A schematic representation of an E-mode shear pattern (left) and a B-mode shear pattern (right) around a point in the sky.

2.2.3. Cosmic shear – Gravitational lensing by the large-scale structure

Our previous efforts have been based on the assumption that all the matter between the source and the observer is concentrated in one single thin lens plane. This approximation may be justified in strong lensing, as most of the lensing signal stems from, for example, a massive galaxy cluster. In our work, we focus on *cosmic shear*, which describes the lensing effect from the LSS in the Universe. There, the lensing effect stems from the entire matter distribution between the source and observer, meaning that the thin lens approximation is no longer valid. However, we can still apply the formalism we developed in the last section to this problem. The exact calculations are quite involved (see Bartelmann and Schneider, 2001; Bartelmann, 2010; Schneider, 2012), so we present a heavily condensed version here that focuses on the central concept and highlights the used approximations.

It can be shown (see e.g. Seitz et al., 1994) that for a weakly perturbed Universe with gravitational potential $|\Phi| \ll c^2$, the FLRW metric becomes

$$\mathrm{d}s^2 = \left(1 + \frac{2\Phi}{c^2}\right) c^2\,\mathrm{d}t^2 - a^2(t)\left(1 - \frac{2\Phi}{c^2}\right)\left[\mathrm{d}\chi^2 + f_K^2(\chi)\,(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)\right] . \tag{2.97}$$

We can then use the fact that the *optical tidal matrix* $\mathcal{T}$ governs the evolution of the transverse physical separation $\boldsymbol{\xi}$ between two light-rays via $\frac{\mathrm{d}^2\boldsymbol{\xi}}{\mathrm{d}\lambda^2} = \mathcal{T}\boldsymbol{\xi}$. From these, one can obtain the evolution equation for the comoving separation $\boldsymbol{x}(\boldsymbol{\theta}, \chi)$ and, in analogy to Sect. 2.2.1, define a Jacobian for the lens mapping

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, \chi) = \delta_{ij} - \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{f_K(\chi - \chi')\, f_K(\chi')}{f_K(\chi)}\, \Phi_{,ik}(\boldsymbol{x}(\boldsymbol{\theta}, \chi'), \chi')\, \mathcal{A}_{kj}(\boldsymbol{\theta}, \chi') , \tag{2.98}$$

where we have made use of the Einstein sum convention. Unfortunately, it is hard to solve this equation as knowledge of $\mathcal{A}(\boldsymbol{\theta}, \chi)$ requires knowledge of all $\mathcal{A}(\boldsymbol{\theta}, \chi')$ and $\boldsymbol{x}(\boldsymbol{\theta}, \chi')$ for $\chi' < \chi$. However, linearising this equation in terms of Φ yields

$$\mathcal{A}_{ij}(\boldsymbol{\theta}, \chi) = \delta_{ij} - \frac{2}{c^2}\int_0^\chi \mathrm{d}\chi' \frac{f_K(\chi - \chi')\, f_K(\chi')}{f_K(\chi)}\, \Phi_{,ij}(f_K(\chi')\,\boldsymbol{\theta}, \chi') . \tag{2.99}$$

We see that, to linear order, the Jacobian only depends on the tidal gravitational field along the unperturbed light-ray $f_K(\chi)\,\boldsymbol{\theta}$. This is called the *Born approximation* and it leads to the fact that the Jacobian $\mathcal{A}_{ij}$ is always symmetric.

We can then, again in analogy to Sect. 2.2.1, define a deflection potential Ψ_{ij} that satisfies $\mathcal{A}_{ij} = \delta_{ij} - \Psi_{,ij}$:

$$\Psi(\boldsymbol{\theta}, \chi) = \frac{2}{c^2} \int_0^\chi \mathrm{d}\chi' \, \frac{f_K(\chi - \chi')\, f_K(\chi')}{f_K(\chi)} \, \Phi(f_K(\chi')\,\boldsymbol{\theta}, \chi') \; . \tag{2.100}$$

Finally, we calculate $\kappa = \nabla^2\Psi/2$ by utilising the Poisson equation (2.30) to be

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{3H_0^2\Omega_\mathrm{m}}{2c^2} \int_0^\chi \mathrm{d}\chi' \, \frac{f_K(\chi - \chi')\, f_K(\chi')}{f_K(\chi)} \, \frac{\delta(f_K(\chi')\,\boldsymbol{\theta}, \chi')}{a(\chi')} \; . \tag{2.101}$$

The above equation describes the convergence κ at a fixed comoving distance χ. While this quantity is an important output of N-body simulations, it can not be observed. Whenever we measure the shear for a sample of galaxies, they follow a certain probability distribution in comoving distance $p_\chi(\chi)\,\mathrm{d}\chi$. For this galaxy sample, we can define the weighted lens efficiency as

$$g(\chi) = \int_\chi^{\chi_\mathrm{H}} \mathrm{d}\chi' \, p_\chi(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')} \; , \tag{2.102}$$

where χ_H is the comoving horizon distance. We can then calculate the effective convergence $\kappa(\boldsymbol{\theta})$ as

$$\kappa(\boldsymbol{\theta}) = \int \mathrm{d}\chi \; p_\chi(\chi)\, \kappa(\boldsymbol{\theta}, \chi) = \frac{3H_0^2\,\Omega_\mathrm{m}}{2c^2} \int_0^{\chi_\mathrm{H}} \mathrm{d}\chi \; g(\chi)\, f_K(\chi) \frac{\delta(f_K(\chi)\,\boldsymbol{\theta}, \chi)}{a(\chi)} \; . \tag{2.103}$$

In practice, probability distributions for galaxies are rarely given as a function of comoving distance, $p_\chi(\chi)\,\mathrm{d}\chi$, but much rather as a function of redshift $p_z(z)\,\mathrm{d}z$. To numerically calculate the effective convergence $\kappa(\boldsymbol{\theta})$ and subsequent quantities that we introduce later, it normally makes the most sense to write the integrals in Eqs. (2.103) and (2.102) as an integral in redshift by substituting $p_\chi(\chi)\,\mathrm{d}\chi = p_z(z)\,\mathrm{d}z$, $a(\chi) = 1/(1+z)$ and $\mathrm{d}\chi = \frac{c}{H(z)}\mathrm{d}z$.

Of course, we do not know our Universe's distribution of density fluctuations δ. What we are interested in, however, are its statistical properties in terms of the power- and bispectrum. For a given source redshift distribution $p_z(z)\,\mathrm{d}z$, we relate the matter power- and bispectrum to the respective convergence power- and bispectrum in the following sections.

2.2.4. Two-point statistics in cosmic shear

Two-point statistics have been the main choice in analyses of cosmic shear (Schneider et al., 1998; Troxel et al., 2018; Hildebrandt et al., 2017; Hikage et al., 2019; Asgari et al., 2020; Hildebrandt et al., 2020). As discussed in Sect. 2.1.2.4, they manage to capture the entire information content of Gaussian random fields. As the field of primordial density fluctuations is (very close to) Gaussian, two-point statistics manage to capture a large part of the information contained in the LSS. This section outlines the two-point cosmic shear statistics relevant to our work and discuss their respective advantages and disadvantages.

2.2.4.1. Convergence power spectrum

Technically, the effective convergence is a function on the unit sphere $\kappa : S^2 \to \mathbb{R}$. This means that the basis for the convergence power spectrum P_κ are spherical harmonics Y_{lm}, Y^*_{lm}. To ease calculations, we instead assume that the convergence is a function in 2-dimensional euclidean space $\kappa : \mathbb{R}^2 \to \mathbb{R}$, which is called the *flat sky approximation*.

Given two comoving distance distributions of sources, $p_{1,2}(\chi)\,\mathrm{d}\chi$, we can calculate their respective lensing efficiencies $g_{1,2}(\chi)$, given a cosmological model. To then compute their convergence power spectrum $P^{12}_\kappa(\ell)$, we calculate

$$\left\langle \tilde{\kappa}_1(\boldsymbol{\ell})\tilde{\kappa}_2^*(\boldsymbol{\ell'})\right\rangle = \frac{9H_0^4\Omega_\mathrm{m}^2}{4c^4}\int_0^{\chi_\mathrm{h}}\mathrm{d}\chi\int_0^{\chi_\mathrm{h}}\mathrm{d}\chi'\,\frac{g_1(\chi)\,g_2(\chi')}{a^2(\chi)}\left\langle\tilde{\delta}\left(\frac{\boldsymbol{\ell}}{f_K(\chi)},\chi\right)\tilde{\delta}^*\left(\frac{\boldsymbol{\ell}}{f_K(\chi')},\chi'\right)\right\rangle\,. \tag{2.104}$$

To ease calculations, we assume that the functions $g_i(\chi)/a(\chi)$ are constant over the range where the correlation between density contrasts at different comoving distances $\left\langle\tilde{\delta}\left(\frac{\boldsymbol{\ell}}{f_K(\chi)},\chi\right)\tilde{\delta}^*\left(\frac{\boldsymbol{\ell}}{f_K(\chi')},\chi'\right)\right\rangle$ is nonzero. Under this assumption, also called *Limber approximation* (Limber, 1954; Peebles, 1980), one can derive (see the Appendix of Kaiser, 1992, 1998; Schneider et al., 1998)

$$\begin{aligned}\left\langle \tilde{\kappa}_1(\boldsymbol{\ell})\tilde{\kappa}_2^*(\boldsymbol{\ell'})\right\rangle &= (2\pi)^2\delta_\mathrm{D}(\boldsymbol{\ell}-\boldsymbol{\ell'})\,\frac{9H_0^4\Omega_\mathrm{m}^2}{4c^4}\int_0^{\chi_\mathrm{h}}\mathrm{d}\chi\frac{g_1(\chi)g_2(\chi)}{a^2(\chi)}\,P_\delta\left(\frac{\ell}{f_K(\chi)},\chi\right)\\ &\equiv (2\pi)^2\,\delta_\mathrm{D}(\boldsymbol{\ell}-\boldsymbol{\ell'})\,P^{12}_\kappa(\ell)\,.\end{aligned} \tag{2.105}$$

It is important to note that the mathematically rigorous derivation of this equation is a fair bit more complicated, yet it yields the same result (see e.g. Bartelmann, 2010).

In weak gravitational lensing, we often split the source galaxies into different tomographic bins, depending on their redshift. For example, in KiDS the source galaxies are split into five redshift bins, that all have different redshift distributions $p_i(z)\,\mathrm{d}z$. When analysing that data, we can calculate the auto-spectra P^{ii}_κ of each redshift bin, but also all the cross-spectra P^{ij}_κ between the individual bins. This allows us to track the evolution of structure over cosmic time. An analysis that includes a split of the source galaxies into different redshift bins is called a tomographic analysis.

We can now relate the statistical properties of matter fluctuations, parametrised by their power spectrum $P_\delta(k,z)$, to the statistical properties of the convergence field. While we have reasonably good models for the power spectrum of matter fluctuations, obtaining the convergence spectra from observations is complicated. Even under the assumption that we can reconstruct an unbiased estimate of the convergence field $\kappa(\boldsymbol{\theta})$ (which is certainly not self-evident), the construction of its Fourier transform $\tilde{\kappa}(\boldsymbol{\ell})$ is even more challenging, as the observed field often has a complex geometry and many gaps, for example, due to bright foreground stars. While there have been successful attempts to mitigate these effects, for example, in the pseudo-C_ℓ framework (Hivon et al., 2002; Elsner et al., 2017; Alonso et al., 2019), we can circumvent the problem entirely by relying on statistics that do not require knowledge about the Fourier transform of a reconstructed convergence field.

2.2.4.2. Shear two-point correlation functions

The arguably most natural way to capture the cosmic shear statistics for an ensemble of galaxies is via its correlation functions. Considering a pair of galaxies at positions $\boldsymbol{\theta}$ and

$\boldsymbol{\theta}'$, we can define the products of their tangential shears with respect to each other as

$$\gamma_\mathrm{t}\gamma_\mathrm{t}(|\boldsymbol{\theta}' - \boldsymbol{\theta}|) \equiv \gamma_\mathrm{t}(\boldsymbol{\theta}; \boldsymbol{\theta}' - \boldsymbol{\theta})\gamma_\mathrm{t}(\boldsymbol{\theta}'; \boldsymbol{\theta}' - \boldsymbol{\theta}) \; . \tag{2.106}$$

Likewise, we define $\gamma_\mathrm{t}\gamma_\times$ and $\gamma_\times\gamma_\times$. We can then define the *shear two-point correlation functions* $\xi_\pm$ as

$$\xi_\pm(\theta) = \langle\gamma_\mathrm{t}\gamma_\mathrm{t}(\theta)\rangle \pm \langle\gamma_\times\gamma_\times(\theta)\rangle \; . \tag{2.107}$$

Recalling Eq. (2.90), we can see that the following equations hold:

$$\begin{aligned} \xi_+(\theta) &= \langle\gamma(\boldsymbol{\vartheta})\gamma^*(\boldsymbol{\vartheta} + \boldsymbol{\theta})\rangle \\ \xi_-(\theta) &= \langle\gamma(\boldsymbol{\vartheta})\gamma(\boldsymbol{\vartheta} + \boldsymbol{\theta})\exp\{-4\mathrm{i}\varphi_{\boldsymbol{\theta}}\}\rangle - \mathrm{i}\,\langle\gamma_\mathrm{t}\gamma_\times + \gamma_\times\gamma_\mathrm{t}\rangle \; , \end{aligned} \tag{2.108}$$

where $\varphi_{\boldsymbol{\theta}}$ is the polar angle of $\boldsymbol{\theta}$ and $\langle\gamma_\mathrm{t}\gamma_\times\rangle = 0$ holds due to parity symmetry of the field.[18] In other words, ξ_+ describes the spin-0 quantity $\langle\gamma\gamma^*\rangle$, whereas ξ_- describes the spin-4 quantity $\langle\gamma\gamma\rangle$. From Eq. (2.87) we know

$$\begin{aligned} \langle\tilde{\gamma}(\boldsymbol{\ell})\tilde{\gamma}^*(\boldsymbol{\ell}')\rangle &= (2\pi)^2\delta_\mathrm{D}(\boldsymbol{\ell} - \boldsymbol{\ell}')P_\kappa(|\boldsymbol{\ell}|) \\ \langle\tilde{\gamma}(\boldsymbol{\ell})\tilde{\gamma}(\boldsymbol{\ell}')\rangle &= (2\pi)^2\delta_\mathrm{D}(\boldsymbol{\ell} + \boldsymbol{\ell}')P_\kappa(|\boldsymbol{\ell}|)\exp(4\mathrm{i}\varphi_{\boldsymbol{\ell}-\boldsymbol{\ell}'}) \; . \end{aligned} \tag{2.109}$$

Calculating the Fourier transforms of Eq. (2.108) and inserting (2.109), we get

$$\xi_\pm(\theta) = \frac{1}{2\pi}\int_0^\infty \mathrm{d}\ell\,\ell\,P_\kappa(\ell)J_{0,4}(\ell\theta) \; , \tag{2.110}$$

where the J_i are the i-th order Bessel functions of the first kind.

Likewise, we can also reconstruct the power spectrum from the shear two-point correlation functions: By making use of the orthogonality relation for Bessel functions

$$\int_0^\infty \mathrm{d}\theta\,\theta\,J_i(\ell\theta)J_i(\ell'\theta) = \frac{\delta_\mathrm{D}(\ell - \ell')}{\ell} \; , \tag{2.111}$$

we can show that

$$P_\kappa(\ell) = 2\pi\int_0^\infty \mathrm{d}\theta\,\theta\,\xi_+(\theta)\,J_0(\ell\theta) = 2\pi\int_0^\infty \mathrm{d}\theta\,\theta\,\xi_-(\theta)\,J_4(\ell\theta) \; , \tag{2.112}$$

holds.

The shear two-point correlation functions have the major advantage that they can be measured in an unbiased way from observed galaxy ellipticities ϵ via the estimator

$$\hat{\xi}_\pm(\theta) = \frac{\sum_{i,j} w_i w_j(\epsilon_{\mathrm{t},i}\epsilon_{\mathrm{t},j} \pm \epsilon_{\times,i}\epsilon_{\times,j})}{\sum_{i,j} w_i w_j} \; , \tag{2.113}$$

where the w_i are potential weights, the $\epsilon_{\mathrm{t},\times}$ are tangential and cross ellipticities defined in analogy to Eq. (2.90), and the sum is over all pairs of galaxies that have a separation of θ. Provided the measured ellipticities are an unbiased estimate of the shear γ, and the

[18] We ignore a potential parity-violation (as reported in Minami and Komatsu, 2020; Philcox, 2022) in this work.

detection of galaxies is random[19], this estimator is unbiased, no matter how complex the survey geometry may be.

As the two-point correlation functions can be directly calculated from the matter power spectrum and easily measured in a survey, they are a good candidate for a cosmological parameter analysis. However, in recent years cosmologists have moved away from this statistic for three reasons. The first is that the shear 2pcf do not have a good E- and B-mode decomposition (compare Schneider et al., 2002). Secondly, the filter functions J_0 and J_4 are quite broad and extend to infinity, which means applying hard scale cuts (for example, because one can not model the power spectrum accurately for $\ell > 10^5$) is virtually impossible. The third one is that the likelihood of the shear 2pcf is not Gaussian (see Sellentin and Heavens, 2018), which is a central assumption in cosmological parameter forecasts, and, when violated, can lead to significant biases (Sellentin et al., 2018).

2.2.4.3. Second-order aperture mass statistics

A different measure for cosmic shear was proposed by Schneider et al. (1998) in the form of the aperture mass dispersion $\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle$, defined as

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle (\theta_{\rm ap}) \equiv \left\langle \mathcal{M}_{\rm ap}^2(\boldsymbol{\theta}; \theta_{\rm ap}) \right\rangle_{\boldsymbol{\theta}} \ . \tag{2.114}$$

Given a cosmological model for the power spectrum, we can easily calculate that statistic. Starting from Eq. (2.91), we derive

$$\left\langle \mathcal{M}_{\rm ap}^2 \right\rangle (\theta_{\rm ap}) = \int_0^\infty \mathrm{d}\ell \, \tilde{u}^2(\ell\theta_{\rm ap}) \, P_\kappa(\ell) \ , \tag{2.115}$$

with $U_{\theta_{\rm ap}}(\theta) = {\theta_{\rm ap}}^{-2} u(\theta/\theta_{\rm ap})$, and $\tilde{u}$ its Fourier transform. For our choice of filter functions (Eq. 2.94), we have

$$\tilde{u}(\eta) = \frac{\eta^2}{2} \exp\left(-\frac{\eta^2}{2}\right) \ . \tag{2.116}$$

Unfortunately, the aperture mass dispersion is difficult to measure in a real survey. While an estimator for the aperture mass can easily be constructed from Eq. (2.92), the filter functions in Eq. (2.94) extend up to infinity and have significant power up to $\theta \sim 4\theta_{\rm ap}$. This means that around any mask or survey boundary, one would have to cut off a strip of $4\theta_{\rm ap}$ width, which leads to a lot of data being thrown out.

Luckily, there is a better way. By writing the correlation functions in the form of Eq. (2.108), one can show that

$$\begin{aligned} \left\langle \mathcal{M}_{\rm ap}^2 \right\rangle (\theta_{\rm ap}) &= \frac{1}{2} \int \mathrm{d}\theta \, \theta \left[\xi_+(\theta) T_+\left(\frac{\theta}{\theta_{\rm ap}}\right) + \xi_-(\theta) T_-\left(\frac{\theta}{\theta_{\rm ap}}\right) \right] \ , \\ \left\langle \mathcal{M}_{\perp}^2 \right\rangle (\theta_{\rm ap}) &= \frac{1}{2} \int \mathrm{d}\theta \, \theta \left[\xi_+(\theta) T_+\left(\frac{\theta}{\theta_{\rm ap}}\right) - \xi_-(\theta) T_-\left(\frac{\theta}{\theta_{\rm ap}}\right) \right] \ , \end{aligned} \tag{2.117}$$

[19] In practice, thwe detected galaxy positions correlate with observational systematics (see e.g. Heydenreich et al., 2020) and foreground matter distributions (Hartlap et al., 2011), which introduces small biases into the measurements.

with analytic filter functions $T_\pm$ (Schneider et al., 2002; Crittenden et al., 2002; Jarvis et al., 2004). The term $\langle \mathcal{M}_\mathrm{ap}\mathcal{M}_\perp \rangle$ vanishes for any parity-invariant field, as does any B-mode contribution of odd power $\left\langle \mathcal{M}_\mathrm{ap}^n \mathcal{M}_\perp^{2m+1} \right\rangle$ (Schneider, 2003). In principle, the aperture mass dispersion is a second-order statistic that provides a straight-forward E- and B-mode decomposition and is obtained from the power spectrum with a highly localised filter function $\tilde{u}$, meaning that scale-cuts are easy to apply. Furthermore, this statistic is in principle straightforward to obtain from a survey using the shear 2pcf and Eq. (2.117). Unfortunately, the required integration extends from zero up to infinity and thus requires knowledge of the correlation functions $\xi_\pm$ on all scales, specifically down to a separation of 0. This is not possible due to the finite extent of galaxy images (Van Waerbeke et al., 2000), and this imperfect knowledge of the correlation function leads to a mixing of E- and B-modes (Kilbinger et al., 2006).

In recent years, COSEBIs have shown to be an ideal real-space statistic for cosmic shear owing to their E- and B-mode decomposition on a finite θ-inteval (Schneider et al., 2010, 2022; Asgari and Heymans, 2019; Giblin et al., 2021; van den Busch et al., 2022, and others), however, we do not use them for this work.

2.2.5. Third-order shear statistics

We now introduce the equivalent third-order shear statistics used in this thesis. In contrast to second-order shear statistics, they have not yet seen widespread use. A cosmological parameter analysis in a weak lensing survey with third-order shear statistics has been performed once by Fu et al. (2014). However, for current-generation surveys, our models have not been validated to the required accuracy. There is a lot of ongoing effort to optimise the measurement of third-order shear statistics (Secco et al., 2022a), to develop models for nuisance parameters (Pyne and Joachimi, 2021; Pyne et al., 2022; Halder and Barreira, 2022), and to develop new summary statistics (Halder et al., 2021). In this section, we briefly introduce the third-order shear statistics that are relevant for this thesis, namely the shear three-point correlation functions (3pcf) and the third-order aperture mass statistics.

2.2.5.1. The convergence bispectrum

In Sect. 2.2.4.1, we have defined the convergence power spectrum $P_\kappa^{12}(\ell)$ for two comoving distance distributions $p_{1,2}(\chi)\,\mathrm{d}\chi$ of source galaxies. In analogy to that section, for three comoving distance distributions $p_{1,2,3}(\chi)\,\mathrm{d}\chi$, we define the convergence bispectrum as

$$(2\pi)^2\,\delta_\mathrm{D}(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3)\,B_\kappa^{123}(\ell_1,\ell_2,\ell_3) = \langle \tilde{\kappa}_1(\boldsymbol{\ell}_1)\tilde{\kappa}_2(\boldsymbol{\ell}_2)\tilde{\kappa}_3(\boldsymbol{\ell}_3)\rangle \ . \tag{2.118}$$

If we again make use of the flat-sky and Limber approximations, an analogous calculation to the one in Sect. 2.2.4.1 yields (see Bernardeau et al., 1997; Schneider et al., 1998)

$$B_\kappa^{123}(\ell_1,\ell_2,\ell_3) = \frac{27H_0^6\,\Omega_\mathrm{m}^3}{8c^6}\int \mathrm{d}\chi\,\frac{g_1(\chi)\,g_2(\chi)\,g_3(\chi)}{f_K(\chi)\,a^3(\chi)}B_\delta\left(\frac{\ell_1}{f_K(\chi)},\frac{\ell_2}{f_K(\chi)},\frac{\ell_3}{f_K(\chi)};z(\chi)\right) . \tag{2.119}$$

We note that we can compute the convergence bispectra for arbitrary triplets of source distributions to perform a tomographic analysis. However, for this work, we only compute

auto-spectra for third-order shear statistics, meaning that all comoving distance distributions of sources are the same ($p_1(\chi)\,\mathrm{d}\chi = p_2(\chi)\,\mathrm{d}\chi = p_3(\chi)\,\mathrm{d}\chi$). In these cases, we drop the indices ^{12} and ^{123} of the power- and bispectrum.

Some sources (e.g. van Waerbeke et al., 1999; Schneider et al., 2005) introduce a different definition for the convergence bispectrum; they define a bispectrum $\mathcal{B}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$ as

$$\langle\tilde{\kappa}(\boldsymbol{\ell}_1)\tilde{\kappa}(\boldsymbol{\ell}_2)\tilde{\kappa}(\boldsymbol{\ell}_3)\rangle = (2\pi)^2[\mathcal{B}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) + \mathcal{B}(\boldsymbol{\ell}_2, \boldsymbol{\ell}_3) + \mathcal{B}(\boldsymbol{\ell}_3, \boldsymbol{\ell}_1)]\delta_\mathrm{D}(\boldsymbol{\ell_1} + \boldsymbol{\ell_2} + \boldsymbol{\ell_3})\ . \tag{2.120}$$

As the bispectrum is invariant to permutations of its arguments for auto-spectra, the only difference from our definition in Eq. (2.51) is a factor of three.

It is often convenient to parametrise the bispectrum in terms of its sidelengths and the angle between the ℓ-vectors:

$$b(\ell_1, \ell_2, \varphi) = B_\kappa(\ell_1, \ell_2, \ell_3) = 3\mathcal{B}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)\ , \tag{2.121}$$

with $\ell_3 = \sqrt{\ell_1^2 + \ell_2^2 + 2\ell_1\ell_2\cos\varphi}$.

2.2.5.2. Shear three-point correlation functions

Like the bispectrum forms the natural extension to the power spectrum, shear two-point correlation functions can be extended by considering correlations between triplets of galaxies. Given a triple of galaxies at positions $\boldsymbol{X}_{1,2,3}$ and the triangle centre $\boldsymbol{X}_\mathrm{c}$, we can again project the shear values to their tangential- and cross-components with respect to the triangle centre

$$\gamma(\boldsymbol{X}_i) = \gamma_\mathrm{t}(\boldsymbol{X}_i; \boldsymbol{X}_i - \boldsymbol{X}_\mathrm{c}) + \mathrm{i}\gamma_\times(\boldsymbol{X}_i; \boldsymbol{X}_i - \boldsymbol{X}_\mathrm{c})\ . \tag{2.122}$$

The four *natural components* of the shear three-point correlation functions $\Gamma^{(i)}$ were constructed to be invariant under rotations of the triangle; they can then be defined as (Schneider and Lombardi, 2003)

$$\begin{aligned}\Gamma^{(0)} &= \langle\gamma(\boldsymbol{X}_1)\gamma(\boldsymbol{X}_2)\gamma(\boldsymbol{X}_3)\rangle\,, \quad \Gamma^{(1)} = \langle\gamma^*(\boldsymbol{X}_1)\gamma(\boldsymbol{X}_2)\gamma(\boldsymbol{X}_3)\rangle\,,\\ \Gamma^{(2)} &= \langle\gamma(\boldsymbol{X}_1)\gamma^*(\boldsymbol{X}_2)\gamma(\boldsymbol{X}_3)\rangle\,, \quad \Gamma^{(3)} = \langle\gamma(\boldsymbol{X}_1)\gamma(\boldsymbol{X}_2)\gamma^*(\boldsymbol{X}_3)\rangle\ .\end{aligned} \tag{2.123}$$

Our shear 3pcf thus have one spin-6 function, $\Gamma^{(0)}$, and three spin-2 functions $\Gamma^{(1,2,3)}$. We note that the argument for each of these shear three-point correlation functions is a triangle of galaxies, such as depicted in Fig. 2.8. For now, we want to parametrize a triangle by its three side-lengths x_1, x_2 and x_3, oriented in the counter-clockwise direction.

It is important to note that the centre of a triangle is not uniquely defined. Apart from the orthocenter O and centroid C (compare Fig. 2.8), one could also, for example, take the incenter (the centre of the circle that has the three triangle sides $\boldsymbol{x}_i$ as a tangent), or the outcenter (the centre of the circle passing through the three vertices of the triangle), as can be seen in Schneider and Lombardi (2003, Fig. 2,3,4 and 5). An advantageous property of the natural components is that they remain invariant under the selection of a triangle centre, up to multiplication with a phase factor. In particular, the moduli $|\Gamma^{(i)}|$ are invariant under the choice of triangle centre.

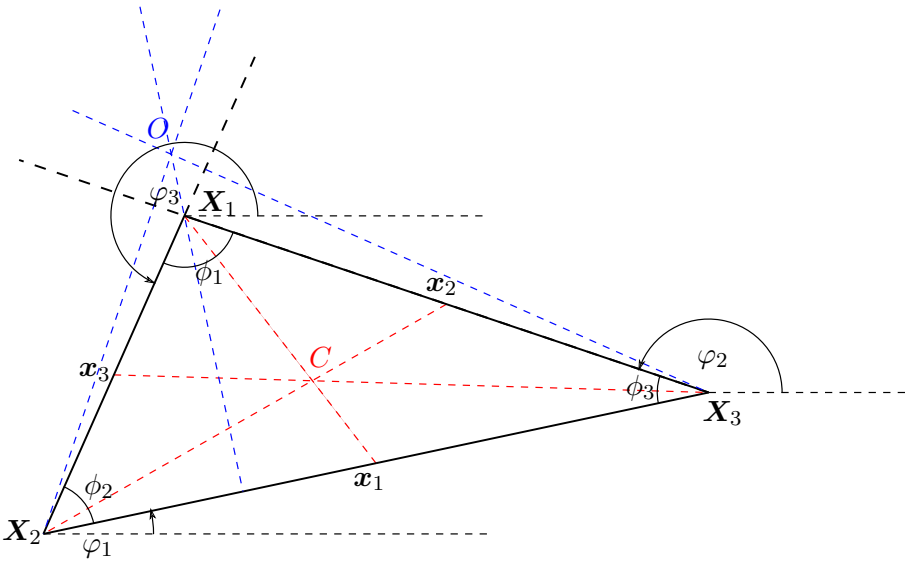


Figure 2.8. – An example triangle with our used notations. The triangle sides $\boldsymbol{x}_i$ are defined as $\boldsymbol{x}_1 = \boldsymbol{X}_3 - \boldsymbol{X}_2$ and cyclic permutations thereof. They take the angles φ_i with respect to the x-axis. The interior angles of the triangle are denoted by ϕ_i. We also show the orthocenter O, the intersection of the three altitudes of the triangle, and its centroid C, the intersection of its three medians (or its centre of mass). Unless otherwise specified, we assume that a triangle is oriented in the counter-clockwise direction. From Heydenreich et al. (2022b).

The shear components transform under a rotation of ζ as $\gamma' = \gamma \exp(-2\mathrm{i}\zeta)$. For example, when switching from the orthocenter to the centroid, the rotations ζ_i for the three shear components γ_i can be determined as (compare Schneider and Lombardi, 2003)

$$\cos(2\zeta_3) = \frac{(x_2^2 - x_1^2)^2 - 4\,x_1^2\,x_2^2\,\sin^2\phi_3}{4\,h_3^2\,x_3^2}\,, \quad \sin(2\zeta_3) = \frac{(x_2^2 - x_1^2)\,x_1\,x_2\,\sin\phi_3}{h_3^2\,x_3^2}\,, \tag{2.124}$$

and cyclic permutations thereof. Here, $h_3 = \frac{1}{2}\sqrt{2\,x_1^2\,x_2^2 - x_3^2}$ is the altitude of the triangle. We can then convert

$$\begin{aligned}
\Gamma^{(0)}_{(\mathrm{cen})} &= \exp\left[-2\mathrm{i}(\zeta_1 + \zeta_2 + \zeta_3)\right]\Gamma^{(0)}_{(\mathrm{o})} \quad \Gamma^{(1)}_{(\mathrm{cen})} = \exp\left[-2\mathrm{i}(-\zeta_1 + \zeta_2 + \zeta_3)\right]\Gamma^{(1)}_{(\mathrm{o})} \\
\Gamma^{(2)}_{(\mathrm{cen})} &= \exp\left[-2\mathrm{i}(\zeta_1 - \zeta_2 + \zeta_3)\right]\Gamma^{(2)}_{(\mathrm{o})} \quad \Gamma^{(3)}_{(\mathrm{cen})} = \exp\left[-2\mathrm{i}(\zeta_1 + \zeta_2 - \zeta_3)\right]\Gamma^{(3)}_{(\mathrm{o})}\,.
\end{aligned} \tag{2.125}$$

Inspecting Eq. (2.123), we can see that the natural components transform into each other under cyclic permutation of indices, in particular

$$\Gamma^{(0)}(x_1, x_2, x_3) = \Gamma^{(0)}(x_3, x_1, x_2), \quad \Gamma^{(1)}(x_1, x_2, x_3) = \Gamma^{(2)}(x_3, x_1, x_2) = \Gamma^{(3)}(x_2, x_3, x_1)\,. \tag{2.126}$$

An anti-cyclic permutation of indices corresponds to a parity transformation, i.e. the new triangle is oriented in a clockwise direction. As, under a parity transformation, $\gamma_\mathrm{t}' = \gamma_\mathrm{t}$ and $\gamma_\times' = -\gamma_\times$ holds, an anti-cyclic permutation of indices leads to a complex conjugation of its corresponding correlation function (in addition to potential transformations according to Eq. 2.126).

The main advantage of the 3pcf is their easy measurement. In analogy to Eq. (2.113), we can construct an estimator

$$\widehat{\Gamma}^{(0)} = \frac{\sum_{ijk} w_i\, w_j\, w_k\, \epsilon_i\, \epsilon_j\, \epsilon_k}{\sum_{ijk} w_i\, w_j\, w_k} \,, \tag{2.127}$$

where the sum extends over all triplets of galaxies whose configuration falls in the respective bin of the estimator. Analogue estimators can be built for $\Gamma^{(1,2,3)}$.

The shear 3pcf can be directly modelled from the matter bispectrum: Taking the orthocenter of the triangle, we observe that the projection direction at point $\boldsymbol{X}_i$ is orthogonal to the opposing triangle side $\boldsymbol{x}_i$. We can thus write the shear projected to the orthocenter $\gamma^{(\mathrm{o})}$ as a rotation of the cartesian shear γ_c via (compare Fig. 2.8)

$$\gamma^{(\mathrm{o})}(\boldsymbol{X}_i) = \gamma_\mathrm{c}(\boldsymbol{X}_i)\mathrm{e}^{-2\mathrm{i}\varphi_i} \,. \tag{2.128}$$

If we now Fourier transform the shear, we can utilise Eq. (2.87) to rewrite

$$\begin{aligned}\Gamma^{(0)}_{(\mathrm{o})}(x_1,x_2,x_3) = \left\langle \gamma^{(\mathrm{o})}(\boldsymbol{X}_1)\gamma^{(\mathrm{o})}(\boldsymbol{X}_2)\gamma^{(\mathrm{o})}(\boldsymbol{X}_3)\right\rangle = \int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_3}{(2\pi)^2} \\ \times\exp\left[-\mathrm{i}(\boldsymbol{\ell}_1\cdot\boldsymbol{X}_1+\boldsymbol{\ell}_2\cdot\boldsymbol{X}_2+\boldsymbol{\ell}_3\cdot\boldsymbol{X}_3)+2\mathrm{i}\sum_i(\beta_i-\varphi_i)\right]\langle\tilde{\kappa}(\boldsymbol{\ell}_1)\tilde{\kappa}(\boldsymbol{\ell}_2)\tilde{\kappa}(\boldsymbol{\ell}_3)\rangle \,,\end{aligned} \tag{2.129}$$

where the β_i are the polar angles of the $\boldsymbol{\ell}_i$. We can now insert Eq. (2.121) and utilise statistical homogeneity and isotropy, to transform this equation to (compare Eq. 15 of Schneider et al., 2005)[20]

$$\begin{aligned}\Gamma^{(0)}_{(\mathrm{o})}(x_1,x_2,x_3) = \frac{2\pi}{3}\int_0^\infty\frac{\mathrm{d}\ell_1\,\ell_1}{(2\pi)^2}\int_0^\infty\frac{\mathrm{d}\ell_2\,\ell_2}{(2\pi)^2}\int_0^{2\pi}\mathrm{d}\varphi\; b(\ell_1,\ell_2,\varphi)\,\mathrm{e}^{2\mathrm{i}\bar{\beta}}\Big[\mathrm{e}^{\mathrm{i}(\phi_1-\phi_2-6\alpha_3)}J_6(A_3) \\ +\mathrm{e}^{\mathrm{i}(\phi_3-\phi_2-6\alpha_1)}J_6(A_1)+\mathrm{e}^{\mathrm{i}(\phi_3-\phi_1-6\alpha_2)}J_6(A_2)\Big] \,.\end{aligned} \tag{2.130}$$

An analogous calculation yields:

$$\begin{aligned}\Gamma^{(1)}_{(\mathrm{o})}(x_1,x_2,x_3) = \frac{2\pi}{3}\int_0^\infty\frac{\mathrm{d}\ell_1\,\ell_1}{(2\pi)^2}\int_0^\infty\frac{\mathrm{d}\ell_2\,\ell_2}{(2\pi)^2}\int_0^{2\pi}\mathrm{d}\varphi\; b(\ell_1,\ell_2,\varphi)\Big[\mathrm{e}^{\mathrm{i}(\phi_1-\phi_2+2\phi_3)}\mathrm{e}^{2\mathrm{i}(\bar{\beta}-\varphi-\alpha_3)}J_2(A_3) \\ +\mathrm{e}^{\mathrm{i}(\phi_3-\phi_2}\mathrm{e}^{-2\mathrm{i}(\bar{\beta}+\alpha_1)}J_2(A_1)+\mathrm{e}^{\mathrm{i}(\phi_3-\phi_1-2\phi_2)}\mathrm{e}^{2\mathrm{i}(\bar{\beta}+\varphi-\alpha_2)}J_2(A_2)\Big] \,.\end{aligned} \tag{2.131}$$

Here, the ϕ_i are the interior angles of the triangle and the x_i are their side-lengths (compare Fig. 2.8), and the 3pcf assume the orthocenter as the triangle cente. The quantities $A_i, \bar{\beta}$,

[20] We note that the different prefactor stems from the different definitions of the convergence bispectrum that we used.

and α_i are defined via cyclic permutations of

$$\begin{aligned}
A_3 &= \sqrt{(\ell_1 x_2)^2 + (\ell_2 x_1)^2 + x_1 x_2\, \ell_1 \ell_2 \cos(\varphi + \phi_3)}\ , \\
|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2|^2 \cos 2\bar{\beta} &= (\ell_1^2 + \ell_2^2)\cos\varphi + 2\ell_1\ell_2\ , \\
|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2|^2 \sin 2\bar{\beta} &= (\ell_1^2 - \ell_2^2)\sin\varphi\ , \\
A_3 \cos\alpha_3 &= (\ell_1 x_2 + \ell_2 x_1)\, \cos\left(\frac{\varphi + \phi_3}{2}\right)\ , \\
A_3 \sin\alpha_3 &= (\ell_1 x_2 - \ell_2 x_1)\, \sin\left(\frac{\varphi + \phi_3}{2}\right)\ .
\end{aligned} \tag{2.132}$$

2.2.5.3. Third-order aperture mass statistics

Like their second-order counterpart, third-order aperture mass statistics can be directly modelled from the convergence bispectrum via (Schneider et al., 2005)[21]

$$\begin{aligned}
\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle (\theta_1, \theta_2, \theta_3) = \frac{2}{(2\pi)^3} \int_0^\infty \mathrm{d}\ell_1\, \ell_1 \int_0^\infty \mathrm{d}\ell_2\, \ell_2 \int_0^\pi \mathrm{d}\varphi\, \tilde{u}(\theta_1\ell_1)\, \tilde{u}(\theta_2\ell_2) \\
\times\, \tilde{u}\left(\theta_3 \sqrt{\ell_1^2 + \ell_2^2 + 2\ell_1\ell_2\cos\varphi}\right) b(\ell_1, \ell_2, \varphi)\ .
\end{aligned} \tag{2.133}$$

It is also always possible to calculate the third-order aperture mass statistics from the shear 3pcf. The conversion is performed by first calculating the quantities

$$\begin{aligned}
\langle MMM \rangle (\theta_1, \theta_2, \theta_3) &= C_1(\theta_1, \theta_2, \theta_3) \int_0^\infty \mathrm{d}y_1 \int_0^\infty \mathrm{d}y_2 \int_0^{2\pi} \mathrm{d}\psi \\
&\quad \times \Gamma^{(0)}_{\rm (cen)}(y_1, y_2, \psi) F_1(y_1, y_2, \psi; \theta_1, \theta_2, \theta_3)\ , \\
\langle MMM^* \rangle (\theta_1, \theta_2; \theta_3) &= C_2(\theta_1, \theta_2, \theta_3) \int_0^\infty \mathrm{d}y_1 \int_0^\infty \mathrm{d}y_2 \int_0^{2\pi} \mathrm{d}\psi \\
&\quad \times \Gamma^{(3)}_{\rm (cen)}(y_1, y_2, \psi) F_2(y_1, y_2, \psi; \theta_1, \theta_2, \theta_3)\ ,
\end{aligned} \tag{2.134}$$

where $\Gamma^{(i)}_{\rm (cen)}(y_1, y_2, \psi)$ are the 3pcf with the centroid as triangle center, parametrized via their two side lengths y_1, y_2 and the interior angle between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. Furthermore $C_{1,2}$ are prefactors and $F_{1,2}$ are filters that both depend on $\theta_{1,2,3}$. Their definitions can be found in Schneider et al. (2005, Eqs. 63-65,69). Then, the third-order aperture statistics

[21] The careful reader might notice that in Schneider et al. (2005) the number in the nominator is a 3, not a 2. In our case, we get a factor 1/3 due to the different bispectrum definition, and a factor 2 since we utilise the symmetry of the bispectrum and only integrate φ from 0 to π.

can be calculated as linear combinations of these quantities:

$$\begin{aligned}
\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle (\theta_1,\theta_2,\theta_3) = \mathrm{Re}\Big[&\left\langle M^2M^* \right\rangle (\theta_1,\theta_2;\theta_3) + \left\langle M^2M^* \right\rangle (\theta_1,\theta_3;\theta_2) \\
&+ \left\langle M^2M^* \right\rangle (\theta_2,\theta_3;\theta_1) + \left\langle M^3 \right\rangle (\theta_1,\theta_2,\theta_3)\Big]/4 \,, \\
\left\langle \mathcal{M}_{\rm ap}^2\mathcal{M}_\perp \right\rangle (\theta_1,\theta_2;\theta_3) = \mathrm{Im}\Big[&\left\langle M^2M^* \right\rangle (\theta_1,\theta_3;\theta_2) + \left\langle M^2M^* \right\rangle (\theta_2,\theta_3;\theta_1) \\
&- \left\langle M^2M^* \right\rangle (\theta_1,\theta_2;\theta_3) + \left\langle M^3 \right\rangle (\theta_1,\theta_2,\theta_3)\Big]/4 \,, \\
\left\langle \mathcal{M}_{\rm ap}\mathcal{M}_\perp^2 \right\rangle (\theta_1;\theta_2,\theta_3) = \mathrm{Re}\Big[&\left\langle M^2M^* \right\rangle (\theta_1,\theta_2;\theta_3) + \left\langle M^2M^* \right\rangle (\theta_1,\theta_3;\theta_2) \\
&- \left\langle M^2M^* \right\rangle (\theta_2,\theta_3;\theta_1) - \left\langle M^3 \right\rangle (\theta_1,\theta_2,\theta_3)\Big]/4 \,, \\
\left\langle \mathcal{M}_\perp^3 \right\rangle (\theta_1,\theta_2,\theta_3) = \mathrm{Im}\Big[&\left\langle M^2M^* \right\rangle (\theta_1,\theta_2;\theta_3) + \left\langle M^2M^* \right\rangle (\theta_1,\theta_3;\theta_2) \\
&+ \left\langle M^2M^* \right\rangle (\theta_2,\theta_3;\theta_1) - \left\langle M^3 \right\rangle (\theta_1,\theta_2,\theta_3)\Big]/4 \,.
\end{aligned} \tag{2.135}$$

Due to parity symmetry, $\langle \mathcal{M}_\perp^3 \rangle$ and $\left\langle \mathcal{M}_{\rm ap}^2\mathcal{M}_\perp \right\rangle$ are expected to vanish. Luckily, for third-order aperture statistics, the problem of E- and B-mode leakage discussed in Kilbinger et al. (2006) does not appear to be a major problem, at least for equal-scale aperture mass statistics with $\theta_1 = \theta_2 = \theta_3$ (Shi et al., 2014). Unfortunately, the functions $C_{1,2}$ and $F_{1,2}$ are difficult to calculate. Jarvis et al. (2004) managed to derive their expressions for equal-scale aperture mass statistics and the filters from Eq. (2.94). Later, Schneider et al. (2005) managed to derive their expressions for general combinations of aperture radii. For other filters, no solutions for $C_{1,2}$ and $F_{1,2}$ have been published.

This procedure has already been used to measure third-order aperture statistics in real data (Fu et al., 2014; Secco et al., 2022a, App. B in this work).

2.2.6. Challenges for gravitational lensing observations

In the previous sections, we have described the gravitational lensing formalism and how one can use it to constrain our cosmological model by comparing observations to theoretical predictions (we describe in Sect. 2.3.3 how we achieve that). This endeavour is a bit more complicated in practice, as our theoretical predictions and observations are subject to uncertainties and systematic biases. In this section, we want to review the accuracy of our theoretical predictions and briefly introduce the primary observational and astrophysical sources of uncertainty in the gravitational lensing measurements.

2.2.6.1. Accuracy of the power- and bispectrum models

The weak lensing two- and three-point statistics can be directly modelled from the matter power- and bispectrum. This means that our predictions of weak lensing statistics can only be as good as our knowledge of the power- and bispectrum. Unfortunately, since the weak gravitational lensing effect is measured in a light cone originating from the observer, it is always sensitive to small scales at low redshift, which are highly non-linear. The linear power spectrum is quite well-understood, with different modelling algorithms agreeing to

the sub-per cent level. Using perturbation theory, one can also construct models for the power- and bispectrum that are valid in the quasi-linear regime of $k \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$, but weak lensing is sensitive to even smaller scales. This means that all models for the power- and bispectrum we can use have to be calibrated using N-body simulations. At these scales, different non-linear power spectrum models disagree by up to a few per cent, while for the bispectrum, even the best model disagrees by up to 20% with its calibration simulations. In addition to these inherent uncertainties, this also means that our models for the matter power- and bispectrum can only be as accurate as our N-body simulations. While they are currently the best method to understand non-linear gravitational dynamics, they are far from perfect and are subject to their own intrinsic inaccuracies (compare Sect. 2.1.2.6). While Hilbert et al. (2020); Springel et al. (2021) show that, for a set of weak lensing observables, current N-body simulations and corresponding ray-tracing techniques have converged for a large range of scales under the chosen simulation parameters, their accuracy can not be proven definitively.

Additionally, modelling a weak lensing statistic alone is insufficient to constrain our cosmological model. We also need to assess the uncertainty of the measurement; the covariance of our chosen statistic (compare Sect. 2.3.3). For second-order lensing statistics, theoretical models for the covariance have been established and tested (see e.g. Schneider et al., 1998; Joachimi et al., 2008, 2021). For third-order lensing statistics, such covariance models are being developed (Linke et al. in prep.). In principle, it is possible to estimate the covariance from the measured data via the jackknifing method, but the estimated covariance is often noisy. Furthermore, this method disregards the effects of *super-sample covariance*: Modes in the field of density fluctuations that are larger than the survey area can bias the measurements in a way that is not being accounted for in the jackknifing method (see e.g. Takada and Hu, 2013; Joachimi et al., 2021). The easiest but computationally most expensive way to estimate a covariance matrix is from a set of N-body simulations. When the simulations are constructed to mimic the properties of a weak lensing survey (particularly regarding the number density and redshift distribution of galaxies), one can just measure the desired statistic in hundreds of different realisations of that simulation. The disadvantage to that method is mostly its computational expense. When a covariance matrix C is constructed from a finite set of n simulations, then its inverse, the precision matrix C^{-1}, is biased depending on the number of entries d in the data vector (Hartlap et al., 2007). In particular, for $d \geq n$, the covariance becomes singular, and its inverse can not be constructed. There exist methods to remedy that effect (Sellentin and Heavens, 2016; Percival et al., 2022); however, the fact remains that the number of simulations required to estimate a reliable covariance matrix is much larger than the length of the data vector. As an example, let us consider the KiDS, which splits the galaxy population into five tomographic redshift bins.[22] For two-point statistics, we can measure $\xi_\pm$ in ~ 10 angular bins, meaning for each of the 15 possible redshift bin combinations, we get a data vector of length 20, yielding $d = 300$ dimensions for a data vector. Indeed, this is likely an upper estimate; a cosmological parameter estimation with COSEBIs would require at most half as much. For three-point statistics, however, we can take the aperture mass statistics for five different filter radii, yielding 35 possible combinations of aperture radii and 35 possible combinations of tomographic bins. The resulting

[22]The KiDS-Legacy data release will include a sixth tomographic bin, which we ignore for this example.

data vector would have $d = 1225$ dimensions, even for this relatively conservative choice. For comparison, the three-point correlation functions in 10 bins would have 8000 entries per tomographic bin. The number of simulations to calibrate the covariance matrix for a tomographic three-point analysis is thus prohibitively high.

2.2.6.2. Validity of the used approximations

Once we have fixed our models for the matter power- and bispectrum, we apply a set of approximations within the weak lensing formalism, which we want to review here. Overall, the approximations are valid to a high degree, but one has to be cautious when utilising either very large or very small scales in the analysis. In general, the approximations seem to break down faster for higher-order statistics.

Weak-field limit To linearise the field equations of GR, we assume that the gravitational potential is small, meaning $|\Phi|/c^2 \ll 1$. For all methods discussed in this work, this assumption is valid. Even close to the centres of galaxy clusters, the gravitational potential is of the order 10^{-5}, and the majority of the weak lensing effect is generated in areas where the gravitational potential is much smaller. Cuesta-Lazaro et al. (2018) investigated this effect perturbatively and concluded that it is negligible even for next-generation experiments when utilising second-order statistics. Giblin et al. (2017) showed that the effect could become relevant to a few per cent on extremely large angular scales of $\ell \lesssim 30$. However, these are usually not utilised in a weak lensing analysis.

Reduced shear approximation While all theoretical predictions are calculated for the power- and bispectrum of the shear γ, the true observable on the sky is the reduced shear $g = \gamma/(1-\kappa)$. Since $\kappa \ll 1$ holds almost everywhere in the sky, we usually assume $g \approx \gamma$. This approximation is violated especially at small angular scales, but with relatively moderate scale cuts (e.g. measuring the 2pcf only at scales larger than $0\farcm5$, Asgari et al., 2021), this effect is negligible even for next-generation surveys (Deshpande et al., 2020b). The same holds for the weak lensing bispectrum (Deshpande and Kitching, 2020).

Born-approximation and Lens-Lens coupling In order to treat the gravitational lensing signal by the LSS with our developed weak lensing formalism, we assume that light-rays propagate through the Universe along unperturbed paths $\boldsymbol{x} = f_K(\chi)\boldsymbol{\theta}$. Furthermore, we neglect that the distribution of lenses along the line of sight is correlated. Among other things, this leads to the conclusion that the Jacobian of the lens equation $\mathcal{A}_{ij}$ is symmetric and that the shear field is without B-modes. With increasing redshift, this approximation starts to get violated. For second-order effects, this is especially relevant when considering lensing of the CMB at redshift $z \sim 1100$ (Beck et al., 2018); in cosmic shear, this effect is negligible (Cuesta-Lazaro et al., 2018; Petri et al., 2017; Shapiro and Cooray, 2006; Cooray and Hu, 2002). However, this effect will become relevant when using third-order statistics in next-generation experiments like Euclid or LSST (Dodelson and Zhang, 2005; Petri et al., 2017).

Limber-approximation To construct the convergence power- and bispectrum from the corresponding matter spectra, we have assumed that the matter spectra are not correlated between different redshifts. Especially for small ℓ-modes, this is incorrect. This effect is relatively small for the power spectrum, introducing an error of about 2% on scales of $\ell < 10$ compared to the exact calculations. Beyond limber corrections reduce this discrepancy to the sub per cent level on all scales (Kilbinger et al., 2017) without the need to integrate over spherical Bessel functions. This effect is more relevant for the bispectrum and can introduce order-of-magnitude errors for $\ell \ll 60$ (Deshpande and Kitching, 2020). When using third-order weak lensing statistics, we thus need to exclude angular scales larger than 6°.

2.2.6.3. Astrophysical systematics

Baryonic effects To assess the non-linear clustering of matter, we usually utilise N-body simulations to calibrate our power- and bispectrum models. However, these simulations are conducted with collisionless particles, corresponding to a Universe filled with only dark matter. They trace the large-scale distribution of matter in the Universe reasonably well (at least as far as we can tell), but baryonic effects become relevant on small scales. These effects consist primarily of feedback processes from active galactic nuclei and supernovae, which blow baryons out of the potential wells of their host galaxies, smoothing the matter distribution in return. Furthermore, on galaxy scales, baryons cluster much easier than dark matter since they can shed kinetic energy via collisions and electromagnetic interactions. We utilise N-body simulations that include baryons and their interaction processes to model baryonic effects. These processes are significantly more complicated than gravitational self-interaction, so one has to employ a variety of simplifications. Furthermore, accessible simulation volumes are usually very limited due to the complicated nature of baryonic processes. Several different simulation suites, such as the BAHAMAS, (McCarthy et al., 2017), the IllustrisTNG (Pillepich et al., 2018) or the Horizon-AGN (Dubois et al., 2014) include baryonic feedback processes and achieve comparable results (Huang et al., 2019). In general, baryonic effects manifest in a suppression of the power spectrum at intermediate scales and an increase at small scales (compare Fig. 3 of Chisari et al., 2019).

Baryonic effects constitute one of the most considerable systematic uncertainties for current weak lensing analyses. Current mitigation strategies are to either forward-model baryonic effects (as done by KiDS) or to exclude scales at which baryonic effects are believed to significantly contaminate the signal (as done by DES). For third-order statistics, the response approach has been developed by Halder and Barreira (2022).

Intrinsic alignments One of the major assumptions in weak lensing analyses is that the intrinsic ellipticity of a galaxy is randomly oriented. However, we know that the orientation of galaxy ellipticity depends on the tidal gravitational field at their position, which is correlated with the shear signal (for a comprehensive review, see Troxel and Ishak, 2015). In the framework of linear structure formation, the intrinsic alignment effect can be modelled from the linear power spectrum (Mackey et al., 2002). This is called the linear alignment model. Replacing the linear power spectrum with its non-linear counterpart

yields the non-linear alignment model (NLA), which is often utilised today (Bridle and King, 2007).

For next-generation experiments, we will require more sophisticated intrinsic alignment models. For this purpose, the Tidal Alignment + Tidal Torque (TATT)-model introduced by Blazek et al. (2019) is a promising candidate.

Next to baryonic feedback processes, intrinsic alignments are one of the major uncertainties in current weak lensing studies. One of the major advantages of a combined second- and third-order analysis is that they can both be modelled by the NLA model (Pyne et al., 2022). A combined analysis of second- and third-order statistics can constrain the amplitude of the intrinsic alignment effect internally, resulting in a significant improvement in cosmological parameter constraints (Pyne and Joachimi, 2021).

Distribution of source galaxies In all modelling approaches, we assume that the distribution of source galaxies in the sky is random and not correlated with the foreground matter distribution. However, some effects lead to an inhomogeneous source distribution. For example, weak lensing sources exhibit a clustering in redshift, which can introduce B-modes and bias the inferred two-point statistics (Schneider et al., 2002). However, that effect is not relevant for second-order statistics yet. Additionally, foreground matter overdensities (which lead to strong shear of background galaxies) inhibit a larger density of foreground galaxies, which in turn obscure background galaxies and reduce their detection probability (Hartlap et al., 2011). The severity of this effect is being investigated (Genc et al., in prep.). Additionally, the selection function of a survey is usually position-dependent due to varying observational conditions. At the second-order level, this effect will become relevant for next-generation surveys (Heydenreich et al., 2020).

Additionally, the presence of foreground matter distorts its background. This leads to a decrease (increase) of the effective observed area $\mathrm{d}^2\beta$, but to a magnification (demagnification) of the galaxies in this area, allowing for easier detection. This effect has been studied in detail and is believed to become relevant for next-generation experiments (Deshpande et al., 2020a) for second-order statistics. For third-order statistics, Deshpande and Kitching (2020) show that the effects are sub-dominant.

2.2.6.4. Observational systematics

Telescope systematics The angular resolution of telescopes is limited. Even for a perfectly built telescope in a total vacuum, the Rayleigh criterion leads to the fact that the resolution limit θ_{res} scales as $\theta_{\mathrm{res}} \approx 1.22\,\lambda\,D^{-1}$, where λ is the wavelength of the incoming light and D the diameter of the primary mirror. The light of distant sources additionally gets distorted by the atmosphere and any telescope imperfections. The image of a point-source on the telescope is called its PSF. When we observe a galaxy, its intensity distribution is convolved with the PSF of the telescope. If the PSF is known, it is relatively straightforward to account for its effects in the shear measurement. Luckily, it can easily be measured by observing stars, which intrinsically constitute point sources, meaning their images equal the PSF. Unfortunately, the PSF depends on the position on the charge-coupled device (CCD), on the declination of the telescope pointing (due to small

deformations of the mirror) and, in the case of the HST, on time, since varying exposure to sunlight deforms the telescope.

Additional effects include non-linear responses of the CCD to incoming light (Guyonnet et al., 2015; Plazas et al., 2016) and imperfect charge transfer during the read-out of the CCD (Kannawadi et al., 2016). For next-generation experiments, additional systematics become relevant, such as electrons that remain in the CCD from a previous exposure (Geis et al., 2015) or the wavelength-dependence of the PSF.

All these effects need to be carefully considered, but since we can directly measure and test them, they are unlikely to present major obstacles to a weak lensing analysis.

Shear measurement Even assuming we can perfectly image a galaxy, one still has to infer its ellipticity. Several approaches have been developed for this problem. One can either measure the second-order brightness moments of the image (Kaiser et al., 1995; Luppino and Kaiser, 1997; Hoekstra et al., 1998), fit a template galaxy profile to the image (Miller and CFHTLenS Collaboration, 2012) or use machine learning processes to infer shear values (Gruen et al., 2010, and others). In recent years, the METACALIBRATION and METADETECTION frameworks (Huff and Mandelbaum, 2017; Sheldon et al., 2020) have proven to be effective tools that can estimate the response of the shear estimator on the actual survey data (without the need for any image simulations) and thus be used to assign weights to the individual shear measurements.

An accurate shear measurement is crucial for subsequent weak lensing analyses. We parametrise the shear measurement bias m on the measured ellipticities via $\epsilon_{\mathrm{meas}} = \epsilon_{\mathrm{true}}(1+m)$. Since all shear measurement algorithms can be rigorously tested using image simulations, we have a high level of control over this nuisance parameter.

Photometric redshift estimation To perform any cosmic shear analysis, we need to measure the redshifts of tens of millions of extremely faint galaxies. Measuring spectral energy distributions for each galaxy would take a prohibitive amount of observation time. Instead, one opts to measure the galaxy's flux in different colour bands and then infer the 'true' redshift of the galaxy from these colour measurements. This is an imperfect method, but it has been continuously improved over the years.

The main issue in the photometric redshift estimation is that the spectroscopic calibration sample only imperfectly covers the high-dimensional colour space of the photometric images. This has recently been addressed by utilising self-organised maps (Wright et al., 2020), which can be used to discard the part of the data that can not be calibrated with sufficient accuracy.

2.3. Mathematical and statistical methods

2.3.1. Homology

A part of this work concerns cosmological parameter estimation using persistent homology. In this section, we want to briefly establish this concept, focusing on how to relate the performed computations to an intuitive understanding.[23] For an interested reader, we

[23] Unfortunately, a formal introduction to the topic would go beyond the scope of this thesis.

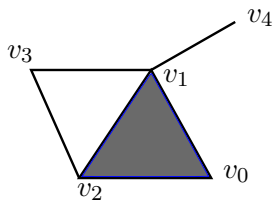


Figure 2.9. – An example of a simplicial complex with one 2-simplex, six 1-simplices and five 0-simplices.

recommend Hatcher (2002) for an introduction into homology (Sect. 2.3.1.1 and 2.3.1.2). A good introduction to the theory of persistent homology can be found in Oudot (2015), whereas Chazal and Michel (2021) provide a summary geared towards its application in data science.

Homology is a concept in the field of algebraic topology, which tries to analyse the topology of objects. In the field of topology, two objects are considered equivalent if they can be 'smoothly' transformed into one another without creating or destroying holes. For example, a mug and a doughnut are the same for all topological purposes. One of the most popular tools to describe the topology of an object X is its homology groups $H_n(X)$. In its essence, for $n \geq 1$ the dimension of the n-th *homology group* $H_n(X)$ of a topological space X is equal to the number of n-dimensional holes, whereas for $n = 0$ the dimension of $H_0(X)$ is the amount of connected components of X.

Before diving into the topic, we want to establish some notations. Let A and B be two vector spaces. For a map $f : A \to B$ that maps an $a \in A$ to $f(a) \in B$, we define the image $\mathrm{Im} f$ as the set of all elements $b \in B$ for which there exists an $a \in A$ such that $f(a) = b$. The kernel $\mathrm{Kern} f$ is defined as the set of all elements a for which $f(a) = 0$ holds.

We also want to introduce the field $\mathbb{F}_2$. It consists of the numbers 0 and 1 and the usual addition and multiplication operators. The rules for addition and multiplication are the same as for integers, apart from the fact that in $\mathbb{F}_2$, $2 = 0$ holds, meaning $1 + 1 = 0$ (or $1 = -1$).

2.3.1.1. Simplicial complexes

Simplicial complexes are arguably the easiest way to define and calculate homology. They present a nice way to relate an intuitive understanding to abstract mathematical concepts. A simplicial complex is built from a set of simplices Δ. Formally, an *n-simplex* Δ^n is defined as the smallest convex set that contains $n + 1$ points $v_0, \ldots, v_n$ whose difference vectors $v_i - v_0$ are linearly independent. For example, a 0-simplex is a dot, a 1-simplex is a line, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron. We want to denote such a simplex by $\Delta^n = [v_0, \ldots, v_n]$. The points v_i are called *vertices* of the simplex.

If $\Delta^n = [v_0, \ldots, v_n]$ is an n-simplex, we denote by $[v_0, \ldots, \hat{v}_i, \ldots, v_n]$ the $(n-1)$ simplex that arises when the vertex v_i gets removed. This $(n-1)$-simplex is called a *face* of Δ^n. As an example, a triangle $[v_0, v_1, v_2]$ has three faces, namely its sides $[v_0, v_1]$, $[v_0, v_2]$ and $[v_1, v_2]$. It is important to note that if a simplex is part of a simplicial complex, all its faces also are part of it.

As an example, Fig. 2.9 depicts a simplicial complex with the 0-simplices (or vertices) v_0, v_1, v_2, v_3, v_4, the 1-simplices $[v_0, v_1], [v_0, v_2], [v_1, v_2], [v_1, v_3], [v_1, v_4], [v_2, v_3]$ and the 2-simplex $[v_0, v_1, v_2]$ (depicted by the gray-shaded region). Note that the triangle spanned by v_1, v_2 and v_3 is empty, meaning that the simplicial complex does *not* include the simplex $[v_1, v_2, v_3]$.[24] Instead, in this position, the complex has a *hole*. This will become quite important later.

2.3.1.2. Simplicial homology

We now want to summarise the topological properties of a simplicial complex with algebraic properties. For a simplicial complex X, we define C_n as the vector space over $\mathbb{F}_2$ that is generated by all n-simplices Δ^n of the complex. In other words, let X be a complex with d n-simplices. The vector corresponding to the first simplex is $(1, 0, 0, \ldots, 0)^T$; the vector corresponding to the second simplex is $(0, 1, 0, \ldots, 0)^T$, and so on. The set of all simplices (or, more precisely, their corresponding vectors) forms a basis of the d-dimensional vector space C_n. In other words, the dimension of C_n already tells us how many n-simplices the complex contains. To describe how a complex is assembled, we construct a boundary map (or boundary matrix) $\partial_n : C_n \to C_{n-1}$. The boundary matrix maps an element from C_n to C_{n-1} via matrix multiplication and is defined as

$$\partial_{n,ij} = \begin{cases} 1 & \Delta_i^{n-1} \text{ is a face of } \Delta_j^n \\ 0 & \text{else} \end{cases} . \tag{2.136}$$

The collection of the vector spaces C_n and respective boundary maps ∂_n,

$$\ldots \longrightarrow C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \longrightarrow \ldots \longrightarrow C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0 ,$$

is called a *chain complex*. One can show that for each n, the combined map $\partial_n \partial_{n+1} : C_{n+1} \to C_{n-1}$ maps everything to the zero-vector, meaning that the image of ∂_{n+1} is a subset of the kernel of ∂_n, $\mathrm{Im}\partial_{n+1} \subseteq \mathrm{Kern}\partial_n$. Given a chain complex with respective boundary maps, we can now always reconstruct the simplicial complex, as we know how many simplices it contains and how they attach to each other. We note that our choice to set $C_n = \mathbb{F}_2^d$ is not necessary; in principle, it is possible to construct chain complexes over many other groups and fields. However, this choice significantly eases the subsequent calculations.

We can now define the n-th homology group $H_n(X)$ as the quotient group $H_n(X) = \mathrm{Kern}\partial_n / \mathrm{Im}\partial_{n+1}$. Since we chose $C_n = \mathbb{F}_2^d$, we can transform this equation to

$$\dim(H_n(X)) = \dim(\mathrm{Kern}\partial_n) - \dim(\mathrm{Im}\partial_{n+1}) . \tag{2.137}$$

From linear algebra, we remember

$$\dim(\mathrm{Kern}\partial_n) + \dim(\mathrm{Im}\partial_n) = \dim(C_n) , \tag{2.138}$$

$$\dim(\mathrm{Im}\partial_n) = \mathrm{rank}(\partial_n) , \tag{2.139}$$

[24] There is no rule preventing $[v_1, v_2, v_3]$ from being a simplex, it was simply a choice made when constructing this example.

where $\mathrm{rank}(\partial_n)$ is the *rank* of the matrix ∂_n (i.e. the number of linearly independent rows or columns of ∂_n). To compute the dimension of $H_n(X)$, we only need to determine the ranks of all the boundary matrices ∂_n. This problem is purely algorithmic and can easily be solved by a computer.

After all these definitions, let us perform an example calculation for the simplicial complex from Fig. 2.9. As the complex has five 0-simplices, six 1-simplices and one 2-simplex, the corresponding chain complex takes the following form:

$$0 \longrightarrow \mathbb{F}_2 \xrightarrow{\partial_2} \mathbb{F}_2^6 \xrightarrow{\partial_1} \mathbb{F}_2^5 \xrightarrow{\partial_0} 0 \,.$$

Our only challenge is constructing the boundary maps ∂_2 and ∂_1 (as ∂_0 maps everything to zero). If we order the simplices as described at the end of Sect. 2.3.1.1, meaning $\Delta_1^1 = [v_0, v_1], \Delta_2^1 = [v_0, v_2], \ldots$, the boundary maps then take the following form (compare Eq. 2.136):

$$\partial_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \,, \quad \partial_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \,. \tag{2.140}$$

Obviously, the rank of ∂_2 is 1. The rank of ∂_1 has to be calculated and is 4 (we can see that the sum of the first four rows equals the last row). Using Eq. (2.137), we can calculate the dimension of homology groups. Since $\dim(\mathrm{Kern}\partial_2) = 0$, $\dim[H_2(X)] = 0$. We have $\dim(\mathrm{Im}\partial_2) = 1$ and, according to Eqs. (2.138) and (2.139)

$$\dim(\mathrm{Kern}\partial_1) = \dim(\mathbb{F}_2^6) - \dim(\mathrm{Im}\partial_1) = 2 \,. \tag{2.141}$$

Following Eq. (2.137), this implies

$$\dim[H_1(X)] = \dim(\mathrm{Kern}\partial_1) - \dim(\mathrm{Im}\partial_2) = \dim(\mathrm{Kern}\partial_1) - \mathrm{rank}(\partial_2) = 2 - 1 = 1 \,. \tag{2.142}$$

At last, we have

$$\dim[H_0(X)] = \dim(\mathrm{Kern}\partial_0) - \dim(\mathrm{Im}\partial_1) = \dim(\mathrm{Kern}\partial_0) - \mathrm{rank}(\partial_1) = 5 - 4 = 1 \,. \tag{2.143}$$

We can thus conclude that the simplicial complex has one hole (the one spanned by the vertices v_1, v_2 and v_3) and one connected component. One could now repeat the calculations and, for example, omit the simplex $[v_0, v_1, v_2]$ (which would yield an additional hole) or add a second 2-simplex $[v_1, v_2, v_3]$ (which would remove the existing hole) and see how the dimensions of the homology groups change.

The fact that the dimension of homology groups corresponds to the number of holes in a simplicial complex is difficult to prove rigorously but can become quite evident when playing around with the example simplicial complex. For example, if we add the second 2-simplex $[v_1, v_2, v_3]$ (and thus close the existing hole), we increase $\dim(\mathrm{Im}\partial_2)$ by one, meaning that we decrease $\dim[H_1(X)]$ by one to $\dim[H_1(X)] = 0$ (compare Eq. 2.142). As another example, we can add an additional vertex v_5 without any connections (meaning

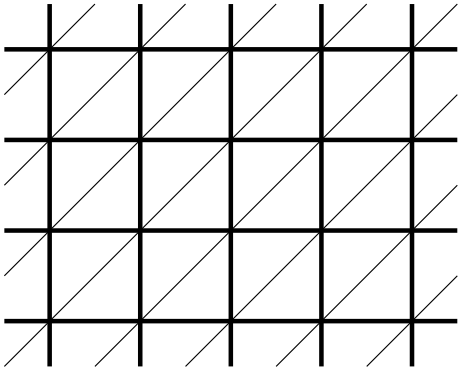


Figure 2.10. – A sketch of how to construct a simplicial complex from a pixel grid. When the thin solid lines are added, we can construct a simplicial complex: Each intersection of lines is a vertex, and each line that connects two intersections is a 1-simplex. Lastly, each triangle is a 2-simplex, meaning that we get two 2-simplices for each pixel.

we add a connected component to the simplex). In that case, $\dim(\mathrm{Kern}\partial_0)$ would increase by 1, and thus $\dim[H_0(X)] = 2$ would hold (compare Eq. 2.143). If we then add a simplex $[v_i, v_5]$, connecting the additional vertex to the existing simplex, we would increase $\dim(\mathrm{Im}\partial_1)$ by one, so that $\dim[H_0(X)] = 1$ holds again.

This means that we have managed to reduce the relatively complex topological question 'How many holes does the space X have?' to a simple algebraic problem (which is the primary goal of all algebraic topology). Of course, for any simplicial complex in less than four dimensions, it would probably be faster and easier to tell how many holes it has by simply looking at it. However, this is nothing that a computer can do. Especially for large simplicial complexes with thousands of simplices, the boundary maps ∂_n are sparse matrices (meaning most of their entries are zero), and computing the rank of a sparse matrix is a well-suited problem for a computer.

2.3.1.3. Persistent homology

Now that we have established the concept of homology, we can introduce *persistent homology*, a method in topological data analysis that quantifies the topological structure of a space on different scales.

Often, this method is applied to point-cloud data, where one constructs a simplicial complex from a set of points. The simplest way of construction is to say that a collection of points $v_0, \ldots, v_n$ form a simplex $[v_0, \ldots, v_n]$ if all pairwise distances $|v_j - v_i|$ are smaller than a scaling parameter r. In this thesis, however, we work with field-level data, for which the concept is applied differently. Everything we discuss below, however, can also be applied to point-cloud data.

As we later apply this method to fields of aperture masses $\mathcal{M}_{\mathrm{ap}}(\boldsymbol{\theta}; \theta_{\mathrm{ap}})$, we explain the concept of persistent homology using this as an example. Given a field of aperture masses, we can easily construct a simplicial complex $X_{\mathcal{M}_{\mathrm{ap}}}$ from that field (compare Fig. 2.10). We then assign the value of each pixel to its corresponding 2-simplices. Afterwards, each n-simplex gets assigned the minimum value of its neighbouring $(n+1)$-simplices. For an arbitrary value t, we can then define a simplicial complex $X_{\mathcal{M}_{\mathrm{ap}}}(t)$ by including all

simplices whose assigned values are smaller than or equal to t. Intuitively, this corresponds to 'cutting off' all values larger than t from the aperture mass field. One can see that for $t < t'$, $X_{\mathcal{M}_{\rm ap}}(t) \subseteq X_{\mathcal{M}_{\rm ap}}(t')$ holds.

When we set $t = \min(\mathcal{M}_{\rm ap}(\boldsymbol{\theta}; \theta_{\rm ap}))$, then the resulting simplicial complex consists of two 2-simplices, five 1-simplices and four 0-simplices, all together forming a square (i.e. the pixel taking the minimum value of the aperture mass field). When we increase t, the simplicial complex grows as more cells are added. Other minima emerge as isolated connected components, which slowly merge (like the purple ellipse in Fig. 2.11). At some point, holes start forming around regions with high aperture mass. Once the threshold surpasses the maximum value of that region, these holes close again (like the red and orange ellipses in Fig. 2.11).

These connected components and holes are *generators* of the homology group $H_n[X_{\mathcal{M}_{\rm ap}}(t)]$. We can track each topological feature by computing the homology group as a function of the scaling parameter. In particular, we can calculate the value at which it first shows up and define that as its birth b. We can also calculate the time it vanishes, its death d. For example, when two connected components merge, one generator of the homology group dies. In that case, we invoke the *elder rule*, that the generator that was born first survives.

When we compute the homology group of this complex as a function of the cut-off value t, we can track the birth and death of all topological features. However, persistent homology offers one additional advantage. Instead of just counting the number of features at each value t, we can calculate the *persistence* (or lifetime) $d - b$ of each feature. This gives us crucial information, as we can distinguish between small fluctuations in the aperture mass and prominent features that significantly protrude from their surroundings (for example, one can see that the red hole in Fig. 2.11 is much more pronounced than the orange one). Other methods from topological data analysis, like Betti numbers or Minkowski functionals, lack this feature and are thus more susceptible to noise.

The fact that each topological feature has a unique generator that can be identified while the filtration value t varies is far from self-evident. However, a proof of that would go beyond the scope of this introduction. For a rigorous proof, we refer the interested reader to Chazal et al. (2016), a summary of that can be found in Chazal and Michel (2021).

We note that the actual persistence computations are executed a bit differently. For field-level data, one usually computes the homology via cubical complexes, where each n-'simplex' is an n-dimensional unit cube $[0, 1]^n$. One can see that the construction of a cubical complex from a pixel grid is a bit more straightforward than that of a simplicial complex and that the resulting computations can be performed faster.

2.3.1.4. Relative homology

One of the main advantages of persistent homology is that it can be calculated in the framework of relative homology, which allows us to compute the homology of a space X with respect to a subspace M, defined as $H_n(X, M)$. Formally, this can be computed via an *exact sequence*

$$\ldots \xrightarrow{\partial_{n+1}} H_n(M) \longrightarrow H_n(X) \longrightarrow H_n(X, M) \xrightarrow{\partial_n} H_{n-1}(M) \longrightarrow \ldots \,,$$

where the image of one map is always exactly equal to the kernel of the next map (in contrast to the chain complex in Sect. 2.3.1.2, where it was just a subset). If we know the dimensions of the homology groups $H_n(X)$ and $H_n(M)$, we can calculate the homology group of $H_n(X, M)$ utilising Eq. (2.138). For our purposes, the relative homology group $H_n(X, M)$ is equivalent to the homology group of the space X/M,[25] which is obtained by taking the space X and collapsing the entire space M to a single point.[26]

For a cosmological parameter analysis, this is incredibly useful: The footprint of a survey never covers the entire sky, and even within the footprint, the presence of bright stars leads to masked regions. This means that the aperture mass field also contains regions over which we have no information that must be masked. Relative homology allows us to calculate the persistent homology of an aperture mass map in the presence of masked regions.

[25] Technically, the *reduced* homology group of X/M.

[26] As an example, if we take a 2-dimensional disk and collapse its entire boundary to a single point, we get a 2-dimensional sphere.

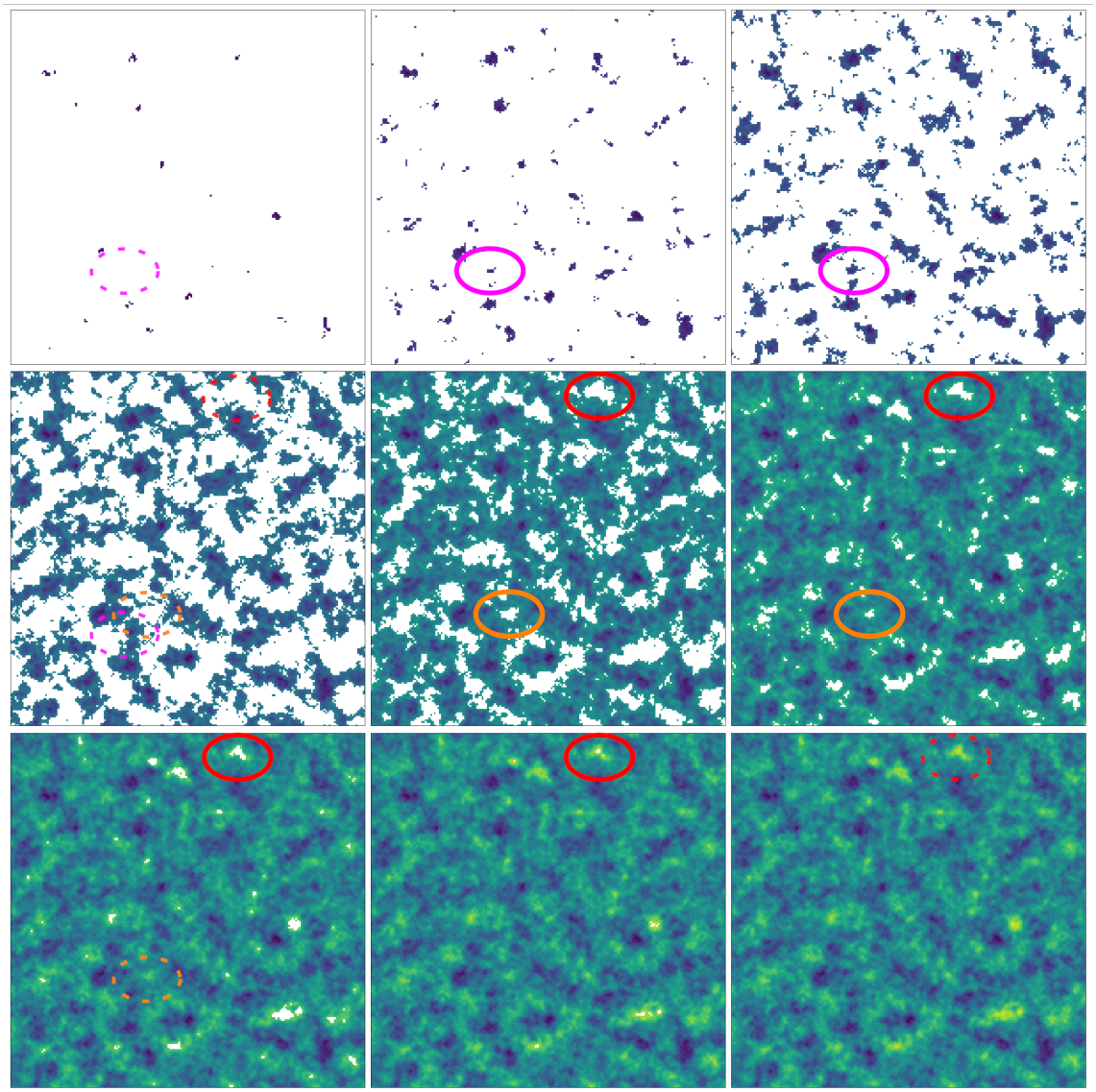


Figure 2.11. – An example aperture mass map $\mathcal{M}_{\rm ap}(\boldsymbol{\theta};\theta_{\rm ap})$ at different values of the cut-off parameter t. The purple ellipses highlight the emergence and disappearance of a connected component; the red and orange ellipses highlight the same for two holes. Figure taken from Heydenreich et al. (2021).

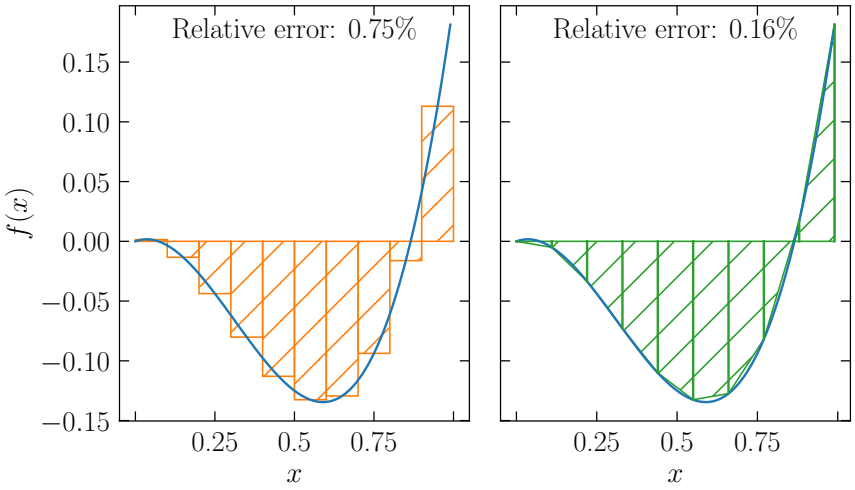


Figure 2.12. – Examples of numerical integration of an arbitrary function $f(x)$ via the rectangle rule (left) and the trapezoidal rule (right).

2.3.2. Numerical integration

Especially for third-order statistics, numerical integration plays a vital role in our work. Numerical integration encompasses methods to determine the integral

$$\int_A \mathrm{d}^n x\, f(\boldsymbol{x}) \approx \sum_i w_i\, f(\boldsymbol{x}_i)\ , \tag{2.144}$$

of a function f over an arbitrary (and possibly infinite) area A and any finite dimension n not via analytic means but rather via a finite amount of function evaluations. Different integration algorithms are simply methods to determine the weights w_i and evaluation points $\boldsymbol{x}_i$ such that the error on the approximation is small for as few function evaluations as possible. Most examples we give in this section assume that the integral is one-dimensional, but they can easily be generalised to higher dimensions. Most integration algorithms are developed to work on the unit cube $[0,1]^n$ or $[-1,1]^n$; an appropriate transformation of variables enables integration over other domains.

Unless otherwise referenced, this section is oriented on Davis et al. (2014), which presents an extensive overview of the different numerical integration methods.

2.3.2.1. Quadrature methods

Quadrature integration methods usually work with high accuracy on various functions, as long as the functions are sufficiently smooth and low-dimensional. The arguably easiest methods for that purpose are the rectangle and the trapezoidal rules (see Fig. 2.12). For the rectangle rule, the evaluation points x_i are evenly spaced throughout the integration domain, and the weights become $w_i \equiv x_1 - x_0$. The function is therefore approximated by a sum of rectangles of width $x_{i+1} - x_i$ and height $f(x_i)$. For the trapezoidal rule, we approximate the function by a sum of trapezoids corresponding to the linear interpolation

between the evaluation points. Although in Fig. 2.12 the trapezoidal rule approximates the integrand much better, both by eye and in the value of the integral, both methods generally perform approximately equally, especially when a larger number of points x_i is sampled.

One of the arguably most popular integration methods is Gaussian quadrature. The general idea is as follows: One constructs the Legendre polynomials, a series of orthonormal polynomials $p_i(x)$ with $\int \mathrm{d}x\, p_i(x)\, p_j(x) = \delta_{ij}$ such that $p_i(x)$ is of degree i. If the function $f(x)$ we want to integrate over is a polynomial of degree $n-1$, it can be represented as a linear combination $f(x) = \sum_{i=0}^{n-1} a_i\, p_i(x)$, where the coefficients a_i can be determined by evaluating the function at n different points. If the integral of the $p_i(x)$ is known, we can calculate the integral of $f(x)$. However, by carefully selecting a specific set of n evaluation points x_i, one can calculate weights w_i such that the relation in Eq. (2.144) is exact for polynomials of degree up to $2n-1$. These evaluation points are closely linked to the orthonormal polynomials $p_i(x)$ and are *not* equally spaced on the integration area. The advantage of Gaussian quadrature is that it gives excellent estimates for relatively few function evaluations. For example, the function in Fig. 2.12 is of degree 3, meaning that only two evaluation points suffice to compute the exact value of its integral. Even for functions that are not polynomials, Gaussian quadrature usually converges towards the true result exponentially as long as the function has no singularities.

Given a function f, we want to determine the value of its integral up to a fixed accuracy ε while using as few evaluation points as possible. This is usually obtained by *adaptive* integration methods. Here, one decides between h-adaptive and p-adaptive integration methods. In a p-adaptive integration method, we integrate the function with our chosen method, double the number of evaluation points, and compute the integral again. Our error estimate is now the difference between the two determined integral values. We subsequently double the number of integration points and re-compute the integral until our error estimate falls below the set accuracy. In a h-adaptive integration method, the integration domain is sub-divided into smaller domains (for example, see Genz and Malik, 1980; Berntsen et al., 1991). The integral and an error estimate are computed on each of these domains. Afterwards, each domain that exceeds the required accuracy is sub-divided into another set of sub-domains, and the procedure is repeated.

The advantage of a h-adaptive quadrature is that convergence is achieved much faster when the function has different properties in different parts of the integration domain (for example, a very isolated peak). With an h-adaptive quadrature, fewer evaluation points are wasted in the 'boring' regions of the integration domain. However, it might be that by pure chance, the error estimate in an 'interesting' region is way smaller than the actual error. In that case, this region would not be sub-divided anymore, and one would potentially miss out on 'interesting' features, thus getting a wrong value for the integral.

The disadvantage of Gaussian quadrature in an adaptive integration algorithm is that the evaluation points x_i are not evenly spaced, meaning that for a $2n$-point Gaussian quadrature, one can not re-use the values $f(x_i)$ from the previously performed n-point Gaussian quadrature. One thus often resorts to Gauss-Kronrod quadrature, which adds $n+1$ evaluation points while re-using the n evaluation points of the Gaussian quadrature.

These adaptive quadrature methods have become the primary method for numerical integration in low dimensions on relatively smooth functions. However, since the integration

needs to be performed in each dimension independently for high-dimensional integrands, the required number of evaluation points quickly becomes too large to handle. To perform these high-dimensional integrations, Monte-Carlo integration methods need to be utilised.

2.3.2.2. Monte Carlo methods

Since quadrature integration methods sample the integration domain on a regular grid, the number of evaluation points increases exponentially with the dimension of the domain. In Monte Carlo methods, one instead samples the integration domain randomly. Given a set of n randomly sampled evaluation points $\boldsymbol{x}_i$, we can set our weights as $w_i = 1/n$ and again estimate the value of the integral using Eq. (2.144). Setting σ^2 as the variance of the function values $f(\boldsymbol{x}_i)$ at the evaluation points, we can define an error estimate as $\varepsilon \approx \sigma/\sqrt{n}$. As one can see, this does not depend on the number of dimensions of the integrand, meaning that convergence can be achieved in a reasonable time by simply increasing the number of evaluation points, even for high-dimensional integrations.

For Monte Carlo methods, adaptive sub-sampling of the integration domain, like in the h-adaptive cubature, is often implemented. If the general shape of the integrand is known, one can also profit from implementing importance sampling methods that sample the evaluation points $\boldsymbol{x}_i$ from a non-uniform probability distribution to place more emphasis on 'interesting' regions of the integrand. As an example, one can determine a coarse histogram of the function f using relatively few evaluation points and then use this approximate reconstruction of f to define a probability distribution for importance sampling. One can then iteratively repeat this procedure, which is done, for example, in the VEGAS integration algorithm (Lepage, 1978).

2.3.2.3. Special-purpose integrators

In the previous subsections, we have introduced cubature and Monte Carlo methods to solve the problem of numerical integration. These methods work well for smooth, well-behaved functions, where cubature methods are usually faster for low dimensions, and Monte Carlo methods excel for high-dimensional integrations. However, both methods have difficulty integrating oscillating functions. Since positive and negative regions of the function partly cancel out in the integration, achieving convergence is much more difficult.

We want to describe our integrals via the form

$$\int \mathrm{d}x\, f(x)\, g(x) \approx \sum_i w_i f(x_i)\ , \tag{2.145}$$

where g is an oscillating function for which we have an analytic expression.[27] If both the function $g(x)$ and its derivative $\mathrm{d}g/\mathrm{d}x$ are known, Levin (1982, 1996) has developed a method to estimate the integral (2.145) where, for quickly oscillating functions, the required amount of integration points with respect to quadrature rules is reduced by a factor of up to 10^5. However, their integration routine has only been established for finite integrals. In our case, the integration over the Bessel function extends to infinity. While it would probably be possible to truncate the integration domain at some point, a different

[27] For our purposes, g is always a Bessel function of the first kind.

method developed by Ogata (2005) avoids this issue. They develop a quadrature rule that uses the zeros of the Bessel function as nodes and show that

$$\int_0^\infty \mathrm{d}x\ f(x)J_\nu(x) \approx \pi \sum_{k=1}^{\infty} w_{\nu k}\ f\left[\frac{\pi}{h}\psi(h\zeta_{\nu k})\right]\ J_\nu\left[\frac{\pi}{h}\psi(h\zeta_{\nu k})\right]\ \psi'(h\,\zeta_{\nu k})\ , \tag{2.146}$$

holds, where the parameter h can be adjusted to fit the desired integration accuracy, $\zeta_{\nu k}$ is the k-th zero of the Bessel function $J_\nu(\pi x)$, and ψ is an arbitrary function. The weights $w_{\nu k}$ are defined as[28]

$$w_{\nu k} = \frac{Y_\nu(\pi\zeta_{\nu k})}{J_{\nu+1}(\pi\zeta_{\nu k})} = \frac{2}{\pi^2\,\zeta_{\nu k}\,J_{\nu+1}^2(\pi\zeta_{\nu k})}\ , \tag{2.147}$$

where Y_ν is the Bessel function of the second kind. If we now choose our function ψ as

$$\psi(t) = t\tanh\left(\frac{\pi}{2}\sinh t\right)\ , \tag{2.148}$$

we see that $\psi(t)$ quickly converges to t (the convergence is double exponentially). For $k \to \infty$, the quadrature nodes $\frac{\pi}{h}\psi(h\zeta_{\nu k})$ thus converge to $\pi\zeta_{\nu k}$ double exponentially, and $J_\nu(\pi\zeta_{\nu k}) = 0$ holds by definition. Therefore, the right-hand side of Eq. (2.146) can be truncated after a small point of function evaluations. This allows us to calculate the infinite integral in (2.146) with a relatively small number of function evaluations; for our purposes, we require less than 30 quadrature nodes.

2.3.3. Statistical foundations of a cosmological parameter analysis

2.3.3.1. Bayesian statistics

Much of this work is dedicated to performing or preparing cosmological parameter analyses, so in this chapter, we want to detail the basic foundations of such an analysis. For any statistic, let $\boldsymbol{d}$ be the data of an observation (for example, the shear two-point correlation functions $\xi_\pm$ in 10 angular bins). We also want to fix a cosmological model with variable model parameters $\boldsymbol{\pi}$, which can give a prediction $\boldsymbol{x}(\boldsymbol{\pi})$ of the observable for each set of model parameters $\boldsymbol{\pi}$.[29] For now, let us assume that the probability distribution of $\boldsymbol{d}$ is a multivariate Gaussian with a covariance matrix C. Due to the multidimensional central limit theorem (van der Vaart, 1998), that is usually a reasonable assumption. Then the probability of observing the data $\boldsymbol{d}$, given a set of model parameters $\boldsymbol{\pi}$ is

$$P(\boldsymbol{d}|\boldsymbol{\pi}) \propto \exp\left(-\chi^2(\boldsymbol{d},\boldsymbol{\pi})/2\right)\ , \tag{2.149}$$

with

$$\chi^2(\boldsymbol{d},\boldsymbol{\pi}) = [\boldsymbol{d}-\boldsymbol{x}(\boldsymbol{\pi})]^T C^{-1}[\boldsymbol{d}-\boldsymbol{x}(\boldsymbol{\pi})]\ . \tag{2.150}$$

While that is a nice fact, it does not help us to constrain cosmological parameters. Instead of the probability of observing a data vector given a set of model parameters, we

[28] We note that the weights defined in Ogata (2005) differ from the ones denoted here due to a typo in the original paper.

[29] The model parameters $\boldsymbol{\pi}$ can include both cosmological and nuisance parameters.

are interested in the likelihood that the model parameters describe a given observation, $\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{d})$. Luckily, we can relate these two via

$$\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{d}) = P(\boldsymbol{d}|\boldsymbol{\pi}) \; . \tag{2.151}$$

However, the likelihood is not necessarily a probability distribution function (i.e. it can not always be normalised). Furthermore, we can not evaluate the likelihood function for all possible model parameters $\boldsymbol{\pi}$, as they span an infinite area. Normally, we do not want to do that either, as we may have some knowledge which parameter values $\boldsymbol{\pi}$ might be sensible. We can then describe the *posterior probability* of the model describing the data using Bayes' theorem (see e.g. Stuart and Ord, 2009)

$$P(\boldsymbol{\pi}|\boldsymbol{d}) = \frac{\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{d})P(\boldsymbol{\pi})}{P(\boldsymbol{d})} \propto P(\boldsymbol{d}|\boldsymbol{\pi})P(\boldsymbol{\pi}) \; . \tag{2.152}$$

Here, $P(\boldsymbol{\pi})$ is the *prior probability* of the model parameters $\boldsymbol{\pi}$ that can for example be taken from previous experiments. Given an observation, we can thus calculate the posterior for our cosmological parameters via

$$\ln P(\boldsymbol{\pi}|\boldsymbol{d}) = \ln \mathcal{L}(\boldsymbol{\pi}|\boldsymbol{d}) + \ln P(\boldsymbol{\pi}) + C = -\chi^2(\boldsymbol{d}, \boldsymbol{\pi})/2 + \ln P(\boldsymbol{\pi}) + C \; , \tag{2.153}$$

where C is a constant that is determined by $P(\boldsymbol{d})$, but can be safely ignored. In a cosmological parameter analysis, great care has to be taken in choosing the prior $P(\boldsymbol{\pi})$. We usually do not want to set the prior as the result of a previous experiment since we want to verify consistency with said experiment first. The canonic choice is to set the prior as uniform within a certain range,

$$P(\boldsymbol{\pi}) = \begin{cases} \prod_i \frac{1}{b_i - a_i} & \pi_i \in [a_i, b_i] \\ 0 & \text{else} \end{cases} \; . \tag{2.154}$$

However, that seemingly 'blind' prior does carry some significant information. For example, we could choose to parametrise our model as a function of a parameter A, or as a function of the parameter $\ln(A)$. In both cases, our model would yield the same predictions, but a uniform prior in A can yield significantly different results than a uniform prior in $\ln(A)$ (compare Joudaki et al., 2020). Thus, when conducting a parameter analysis using Bayesian statistics, one has to be aware that it is impossible to set a truly uninformative prior. Therefore, great care is required when treating priors in any cosmological parameter analysis.

After we have computed a posterior $P(\boldsymbol{\pi}|\boldsymbol{d})$, we are often interested in the *marginalised posterior* of a certain subset $\{\pi_i\}_{i\in I}$ of parameters (for example, we want to know the one-dimensional marginalised posterior of the parameter S_8). We can achieve that by integrating the posterior over all remaining parameters

$$P(\{\pi_i\}_{i\in I}|\boldsymbol{d}) = \left(\prod_{j\notin I} \int \mathrm{d}\pi_j\right) P(\boldsymbol{\pi}|\boldsymbol{d}) \; . \tag{2.155}$$

In principle, these methods allow us to construct a posterior probability distribution of our parameter space. However, when performing a cosmological parameter analysis, we

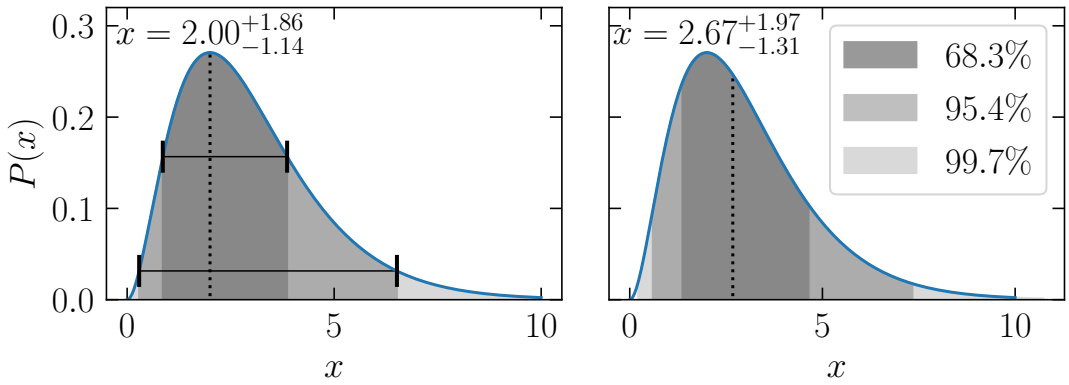


Figure 2.13. – An example of an asymmetric probability distribution. The grey shaded areas indicate the 68.3%, 95.4% and 99.7% credible intervals. In the left panel, they represent the most compact intervals covering the same probabilities as the respective regions within 1, 2 and 3 standard deviations of a Gaussian distribution. In that case, the probability density at the boundaries of each credible interval is the same and the black dotted line represents the maximum value of the posterior. In the right panel, the intervals correspond to the regions that have equal probability in their left and right tails, and the dotted line corresponds to the median of the probability distribution.

mostly want to report constraints on the individual marginalised parameters. There we face the problem that the posterior might be asymmetric. If we want to cite parameter constraints of an asymmetric distribution, we still want to cite a mean (or best fit) value and a 68.3% credible interval (corresponding to the interval within one standard deviation of a Gaussian distribution). There are several ways to do this, as exemplified in Fig. 2.13. When we cite parameter constraints in Chap. 4, we employ the *equal tail* definition of credible intervals, which corresponds to the right panel of Fig. 2.13; in Chap. 5 we use the *highest posterior density intervals*, corresponding to the left panel of Fig. 2.13.

In practice, a covariance matrix C is often obtained using the sample covariance of a (finite) set of observations in N-body simulations. Assuming that the data $\boldsymbol{d}$ follow a multivariate Gaussian probability distribution (as in Eq. 2.149), the sample covariance matrix is an unbiased estimate of the true covariance and follows a Wishart distribution (Anderson, 2003). However, the *precision matrix* C^{-1} becomes biased, and a naive cosmological parameter analysis would yield too optimistic parameter constraints, in particular when the dimension of the data vector is large (Hartlap et al., 2007). One can circumvent this by marginalising the Gaussian likelihood (2.153) over the unknown covariance matrix C, which yields the new likelihood distribution (Sellentin and Heavens, 2016)

$$\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{d}) \propto \left[1 + \frac{\chi^2(\boldsymbol{d}, \boldsymbol{\pi})}{N-1}\right]^{-\frac{N}{2}} , \tag{2.156}$$

where N is the number of independent realisations used to construct the sample covariance matrix.

An extension of this method was recently presented by Percival et al. (2022), who developed a method to match frequentist and Bayesian interpretations of posterior constraints.

They replace the power-law index in Eq. (2.156) by $-m/2$ with

$$m = n_\pi + 2 + \frac{N - 1 + B(n_d - n_\pi)}{1 + B(n_d - n_\pi)}$$
$$B = \frac{N - n_d - 2}{(N - n_d - 1)(N - n_d - 4)} , \tag{2.157}$$

where n_d is the dimension of the data vector $\boldsymbol{d}$ and n_π is the dimension of the cosmological parameter space.

2.3.3.2. Markov-Chain Monte Carlo processes

Now that we have established how to compute the posterior probability distribution for cosmological parameters, a cosmological parameter analysis appears feasible. However, a major problem is the dimensionality of the cosmological parameter space. Assuming we have 14 independent parameters in our cosmological model (when including nuisance parameters, this is rather moderate), evaluating the likelihood function on a uniform grid becomes virtually impossible. Even for an extremely coarse sampling of 10 grid points per dimension, we would have to compute the function 10^{14} times. To circumvent this problem, we rely on the Metropolis-Hastings algorithm.[30] The principle idea is as follows:

We start by computing the likelihood $\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{d})$ at an arbitrary point $\boldsymbol{\pi}$. We then pick a point $\boldsymbol{\pi} + \boldsymbol{\pi}'$ in the neighbourhood of $\boldsymbol{\pi}$. Usually, but not necessarily, $\boldsymbol{\pi}'$ is drawn from a multivariate normal distribution. We then compute the *acceptance ratio*

$$\alpha = \frac{\mathcal{L}(\boldsymbol{\pi} + \boldsymbol{\pi}'|\boldsymbol{d})P(\boldsymbol{\pi} + \boldsymbol{\pi}')}{\mathcal{L}(\boldsymbol{\pi}|\boldsymbol{d})P(\boldsymbol{\pi})} . \tag{2.158}$$

Afterwards, we draw a uniform random number u between 0 and 1. If $u < \alpha$, the point $\boldsymbol{\pi} + \boldsymbol{\pi}'$ gets accepted and we repeat the procedure with $\boldsymbol{\pi} + \boldsymbol{\pi}'$ as the 'new' initial point. Otherwise, the point $\boldsymbol{\pi} + \boldsymbol{\pi}'$ gets rejected, and we repeat the procedure from the start. This algorithm leads to a random walk, and the probability density of accepted points converges to the true posterior (Rosenbluth, 2003). This approach has several advantages: By the nature of the procedure, the interesting regions where the posterior is large are more densely sampled than the uninteresting regions of low posterior probability. Furthermore, marginalisation is extremely easy: Instead of computing a (potentially) high-dimensional integral, one can project the multi-dimensional posterior along the parameter axis by ignoring the marginalised dimensions of the sampled random walk.

Unfortunately, convergence is only ensured after an infinite number of evaluations, so there is the danger that one mistakes a local maximum for the global one and believes that the random walk has already converged. One usually computes multiple random walks with different starting positions in parallel to avoid this issue. Still, convergence can require many evaluations for the likelihood function, which can be prohibitive if it is numerically expensive to compute. In recent years, nested sampling methods such as polychord (Handley et al., 2015) promise to reduce the necessary amount of evaluations to the likelihood function while still recovering the correct shape of the posterior. However,

[30] Alongside Fast Fourier Transforms, the Metropolis algorithm is considered to be one of the ten most important ones of the 20th century (`https://archive.siam.org/pdf/news/637.pdf`).

the success of this endeavour depends on a correct fine-tuning of the hyper-parameters for the sampling method and needs to be adjusted to the survey settings (Lemos et al., 2022).

To transform the resulting point clouds into posterior distributions, we smooth them with a kernel. The size of that kernel has to be chosen appropriately; it should be large enough to smooth out bumps caused by individual points, but not so large that it artificially broadens the posteriors. This allows us to recover a posterior distribution in the low-dimensional projections where we can visualise it. Alternative ways to recover the posterior without invoking a Markov-Chain Monte Carlo (MCMC) run are variational-Bayes solutions (Rizzato and Sellentin, 2022) or likelihood-free inference (Jeffrey et al., 2021).

CHAPTER 3

Used data products

In this chapter we introduce the data products and N-body simulations that were used for this thesis. When talking about weak lensing surveys, we distinguish between Stage-III and Stage-IV surveys (Albrecht et al., 2006). Stage-III surveys are ongoing or recently completed weak lensing surveys such as KiDS, DES or Hyper Surprime-Cam Subaru Strategic Program (HSC SSP, Aihara et al., 2018). Stage-IV represents the next generation of weak lensing surveys, such as LSST or *Euclid* (Laureijs et al., 2011). They are expected to see first light within the next years and will enable us to place meaningful constraints on the dark energy EoS.

3.1. Kilo-Degree Survey

We use the public[1] data from the fourth data release of the KiDS (Kuijken et al., 2015, 2019; de Jong et al., 2015, 2017) carried out by the European Southern Observatory. The KiDS is uniquely optimised for weak lensing studies; the VST-OmegaCAM mounted at the Cassegrain focus of the VST allows for a very uniform and circular PSF, and the overlap with the infrared VISTA Kilo-degree Infrared Galaxy survey (VIKING, Edge et al., 2013) means that galaxies are observed in nine photometric bands, allowing for tight control over potential redshift uncertainties (Hildebrandt et al., 2021), especially at higher redshifts. We use fourth data release of the KiDS (KiDS-1000, Kuijken et al., 2019), which covers roughly $1000\,\mathrm{deg}^2$ on the sky, in Sect. 5.1 and App. B; we also work with the previous KiDS+VIKING-450 (KV450, Wright et al., 2019) data release in Sect. 4.2.

[1] `https://kids.strw.leidenuniv.nl/DR4`

3.2. Dark Energy Survey

In Sect. 4.3, we use the public[2] year-1 data of the Dark Energy Survey (DES-Y1) data presented in Abbott et al. (2018). The main weak lensing data product is a galaxy catalogue, containing positions and ellipticities for tens of millions of galaxies observed with the DECam at the Blanco telescope (Flaugher et al., 2015). We tailor our analysis to use the same galaxies as Troxel et al. (2018, hereafter T+18) and Harnois-Déraps et al. (2021, hereafter HD+21) by picking only galaxies with the FLAGS SELECT, METACAL, and the REDMAGIC filters. In total, we use 26 million galaxies spanning an effective area of about $1321\,\mathrm{deg}^2$.

The galaxy ellipticities $\epsilon_{1/2}$ are inferred using the METACALIBRATION technique (Sheldon and Huff, 2017). This provides each galaxy with a response function S_i that parametrises the sensitivity of the measurement to potential galaxy shear. As in T+18, we provide the method with a prior on the multiplicative shear bias of $m \pm \sigma_m = 0.012 \pm 0.023$. We use the reduced shear approximation and thus assume that the measured ellipticities provide us with an unbiased estimator of the shear γ.

The source sample was split into four tomographic redshift bins according to their photometric redshifts measured with the BPZ method (Benítez, 2000). In contrast to the original DES-Y1 method (Hoyle et al., 2018), we follow HD+21 and determine the redshift distribution of each tomographic bin using the 'DIR' calibration method (Lima et al., 2008). This method was shown to suffer from smaller systematic uncertainties and to be more robust to potential selection effects of the spectroscopic training sample (Joudaki et al., 2020; Gruen and Brimioulle, 2017). The uncertainty on the mean redshift of the source distribution has been determined by bootstrap resampling to be $\sigma_z = 0.008, 0.014, 0.011$ and 0.009 for tomographic bins 1..4, respectively (Joudaki et al., 2020). This method yields excellent agreement between the cosmological parameter constraints of DES-Y1 and KV450 cosmic shear data (Hildebrandt et al., 2020). A comparison between the two methods to determine the redshift distribution shows that the inferred S_8 value of DES-Y1 changes by less than 1σ, which is certainly noticeable but causes no tension between the two different redshift estimation methods.

3.3. Galaxy catalogues from N-body simulations

In addition to the real cosmic shear data described above, we also use several N-body simulations to validate our models, extract covariance matrices, and conduct simulation-based inference. In this section, we detail the different N-body simulations that we used for our analysis in Chapters 4 and 5.

3.3.1. SLICS

The Scinet light cone simulations (SLICS) consist of a set of 927 fully independent N-body simulations conducted in a flat ΛCDM-Universe with $\Omega_\mathrm{m} = 0.2905$, $\Omega_\mathrm{b} = 0.0473$, $h = 0.6898$, $\sigma_8 = 0.826$ and $n_\mathrm{s} = 0.969$. For each simulation, initial conditions were evolved to redshift $z = 0$ with 1536^3 particles in a box of $505\,h^{-1}\,\mathrm{Mpc}$, and subsequently

[2] DES-Y1 catalogues: des.ncsa.illinois.edu/releases/dr1

ray-traced into $10 \times 10\,\mathrm{deg}^2$ light-cones up to $z = 3$, producing between 15 and 28 mass sheets of co-moving thickness equivalent to half the box size.

3.3.1.1. cosmo-SLICS

The cosmo-SLICS are a suite of cosmological N-body simulations in which the matter density Ω_m, the parameter S_8, the Hubble constant h and the parameter for the dark energy equation of state w_0 are sampled at 26 points in a Latin hyper-cube (see Tab. 3.1 at the end of this chapter for the exact list). At each cosmology, a pair of N-body simulations was evolved the same way as the SLICS and subsequently ray-traced multiple times to yield 50 pseudo-independent lines of sight. The initial conditions were chosen such as to suppress most of the sample variance when averaging a statistic over the pair (for more details on the cosmo-SLICS, we refer the reader to Harnois-Déraps et al., 2019).

3.3.1.2. Adjusting simulations to weak lensing surveys

KV450-like We take the KV450 data set (Wright et al., 2019) as an example of a current Stage-III weak lensing survey and create mock data sets with similar properties. Due to the box size of our simulations, the full KV450 survey footprint cannot be fitted onto a single light cone. Instead, we split the survey into 17 tiles following the setup presented in Appendix A3 of Harnois-Déraps et al. (2018) and compute the (simulated) shear signal at the exact positions of the KV450 galaxies, repeating the process for ten light-cones (out of the 50 available) for each cosmo-SLICS pair, and for 126 SLICS realisations. The galaxy redshifts were randomly selected such that the cumulative redshift distribution follows the fiducial "direct calibration" method (DIR) described in Hildebrandt et al. (2020).

DES-Y1-like Similarly, we tailor the SLICS and cosmo-SLICS to match the DES-Y1 survey specifications by splitting the survey into 19 tiles, following the setup in HD+21. Again we compute the simulated shear signal at the positions of the DES-Y1 galaxies and repeat this process for 5 out of the 50 available light-cones for each cosmo-SLICS pair and for 126 SLICS realisations. We randomly select the galaxy redshift distributions so that their cumulative distribution follows the one determined by Joudaki et al. (2020).

***Euclid*-like** We use *Euclid*-like SLICS and cosmo-SLICS in both of the projects described in Chapters 4 and 5. In both cases, we take the full $100\,\mathrm{deg}^2$ maps and randomly distribute source galaxies to achieve a number density of $30\,\mathrm{arcmin}^{-2}$. We sample the source galaxy redshifts between $z = 0$ and $z = 3$ with a probability distribution of

$$n(z) \propto z^2 \exp\left[-\left(\frac{z}{z_0}\right)^\beta\right] , \tag{3.1}$$

with $z_0 = 0.637$ and $\beta = 1.5$. These choices represent our expectations of the performance of the *Euclid*-mission; exact values can not be given as the mission has not been launched yet.

Obtaining realistic galaxy catalogues While we can measure the uncontaminated shear of a source galaxy in simulations, real galaxies suffer from shape noise. Especially for the project in Chap. 4, we are highly sensitive to the survey's exact shape noise properties, so it is crucial to reproduce these in the simulation as accurately as possible. We obtain our mock ellipticities by rotating the ellipticities that are observed in the respective survey (KV450 or DES-Y1) by a random angle and combine the resulting signal ϵ_n with the simulated reduced shear g via Eq. (2.85). The random rotation ensures that the resulting ellipticity ϵ_n carries no cosmic shear signal but traces almost the exact shape noise of the respective survey. This is important as the shape noise of galaxies can depend on position, magnitude, and redshift; local variations of these properties could otherwise bias our statistics.

In total, we compute ten shape-noise realisations for every simulated survey realisation, each using a different random seed in the rotation. This procedure allows us to average out a large part of the fluctuations introduced by the shape noise, improving both our predictions and our estimate of the sample covariance while preserving the data noise levels. For the *Euclid*-like simulations, we draw the shape-noise from a random distribution with zero mean and dispersion $\sigma_\epsilon^2 \equiv \langle|\epsilon_n^2|\rangle \approx 0.3$.

3.3.2. Magneticum simulations

We use the Magneticum simulations[3] (more precisely the Magneticum *Run*-2 and *Run*-2b of Castro et al., 2021) to estimate the impact of baryons on our desired summary statistics. Apart from gravitational interactions, these hydrodynamical simulations include models for radiative cooling, star formation, supernovae and feedback by active galactic nuclei. The simulations were run in volumes of 352 and 640 h^{-1} Mpc, respectively, with the cosmological parameters $\Omega_\mathrm{m} = 0.272$, $h = 0.704$, $\Omega_\mathrm{b} = 0.0451$, $n_\mathrm{s} = 0.963$, and $\sigma_8 = 0.809$. Conveniently, this is quite close to the cosmology of the SLICS, which allows us to use our models for baryonic feedback on the work conducted with these simulations. The feedback determined by the Magneticum simulations is consistent with the one of the BAHAMAS simulations (McCarthy et al., 2017), a completely independent set of hydrodynamical simulations. Both reproduce a number of observations of the LSS, as detailed in Hirschmann et al. (2014); Teklu et al. (2015); Castro et al. (2018).

3.3.3. Millennium simulations

The Millennium simulation (MS) are a set of simulations with a flat ΛCDM cosmology with $h = 0.73$, $\sigma_8 = 0.9$, $\Omega_\mathrm{m} = 0.25$, $\Omega_\mathrm{b} = 0.045$ and $n_s = 1$. They were run with 2160^3 particles in a $500\,h^{-1}\,\mathrm{Mpc}$ box and subsequently ray-traced (Hilbert et al., 2009) to yield shear- and convergence-maps of 64 semi-independent lines of sight with an area of $4 \times 4\,\mathrm{deg}^2$. The shear and convergence maps are provided for each line of sight on a grid of 4096×4096 pixel at 36 different source redshifts.

[3] `www.magneticum.org`

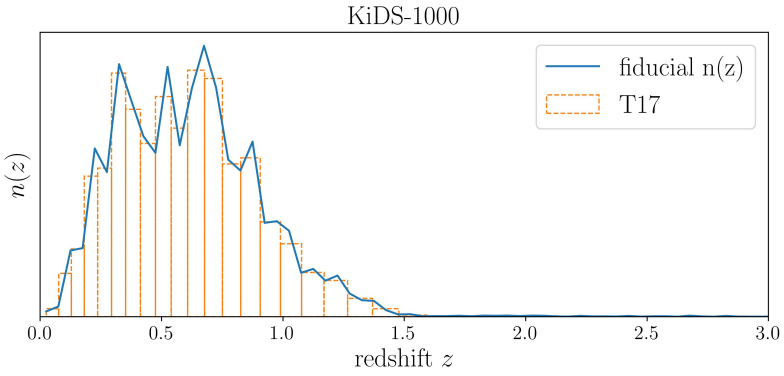


Figure 3.1. – The fiducial redshift distribution of the KiDS-1000 survey (blue) and the respective weights we used for the convergence shells in the T17 simulations (orange). From Heydenreich et al. (2022b).

3.3.4. Full-sky weak lensing convergence maps

We use the public[4] full-sky weak lensing convergence maps (T17, Takahashi et al., 2017) to perform a realistic, curved-sky analysis, mimicking the properties of the KiDS-1000 survey (Giblin et al., 2021). The simulations were constructed from a series of nested cubic boxes with sidelengths $L, 2L, 3L, \ldots$ with $L = 450\,\mathrm{Mpc}/h$. The cubes are placed around a fixed point representing the observer's position. Each box contains 2048^3 dark matter particles that are evolved with the GADGET2 code (Springel et al., 2001) in a flat ΛCDM cosmology with $\Omega_\mathrm{m} = 0.279$, $\Omega_\mathrm{b} = 0.046$, $h = 0.7$, $\sigma_8 = 0.82$, and $n_\mathrm{s} = 0.97$. Then, spherical lens shells with a width of $150\,\mathrm{Mpc}/h$ are constructed and ray-traced using the code GRayTrix[5].

For each realisation, we build realistic convergence maps by computing a weighted average over all 38 convergence shells, where the weight of each shell is determined by the value of the fiducial KiDS-1000 $n(z)$ at the respective redshift (see Fig. 3.1). To include the effect of shape noise, we add a Gaussian random variable with variance

$$\sigma^2_\mathrm{pix} = \frac{\sigma^2_\epsilon}{n_\mathrm{gal} A_\mathrm{pix}}\,, \tag{3.2}$$

where A_pix is the pixel area of the convergence grid. For the galaxy number density and shape noise, we take $n_\mathrm{gal} = 6.17$ and $\sigma_\epsilon = 0.265$, which are the values measured in the KiDS-1000 analysis.

[4] http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/
[5] http://th.nao.ac.jp/MEMBER/hamanatk/GRayTrix/

Table 3.1. – Cosmological parameters of the cosmo-SLICS wCDM simulations. Other fixed parameters are $\Omega_\mathrm{b} = 0.0447$ and $n_s = 0.969$.

	Ω_m	S_8	h	w_0	σ_8	Ω_cdm
00	0.3282	0.6984	0.6766	−1.2376	0.6677	0.2809
01	0.1019	0.7826	0.7104	−1.6154	1.3428	0.0546
02	0.2536	0.6133	0.6238	−1.7698	0.667	0.2063
03	0.1734	0.7284	0.6584	−0.5223	0.9581	0.1261
04	0.3759	0.8986	0.6034	−0.9741	0.8028	0.3286
05	0.4758	0.7618	0.7459	−1.3046	0.6049	0.4285
06	0.1458	0.768	0.8031	−1.4498	1.1017	0.0985
07	0.3099	0.7861	0.694	−1.8784	0.7734	0.2626
08	0.4815	0.6804	0.6374	−0.7737	0.5371	0.4342
09	0.3425	0.7054	0.8006	−1.501	0.6602	0.2952
10	0.5482	0.6375	0.7645	−1.9127	0.4716	0.5009
11	0.2898	0.7218	0.6505	−0.6649	0.7344	0.2425
12	0.4247	0.7511	0.6819	−1.1986	0.6313	0.3774
13	0.3979	0.8476	0.7833	−1.1088	0.736	0.3506
14	0.1691	0.8618	0.789	−1.6903	1.1479	0.1218
15	0.1255	0.6131	0.7567	−0.9878	0.9479	0.0782
16	0.5148	0.8178	0.6691	−1.3812	0.6243	0.4675
17	0.1928	0.8862	0.6285	−0.8564	1.1055	0.1455
18	0.2784	0.65	0.7151	−1.0673	0.6747	0.2311
19	0.2106	0.8759	0.7388	−0.5667	1.0454	0.1633
20	0.443	0.8356	0.6161	−1.7037	0.6876	0.3957
21	0.4062	0.662	0.8129	−1.9866	0.5689	0.3589
22	0.2294	0.8226	0.7706	−0.8602	0.9407	0.1821
23	0.5095	0.7366	0.6988	−0.7164	0.5652	0.4622
24	0.3652	0.6574	0.7271	−1.5414	0.5958	0.3179
fid	0.2905	0.8231	0.6898	−1.0	0.8364	0.2432

CHAPTER 4

Persistent homology in cosmic shear

Persistent homology is a powerful tool from topological data analysis that can be applied to a wide range of data products. It has already seen successful applications in identifying subgroups of cancer (Nicolau et al., 2011), analysing the spread of contagious diseases (Lo and Park, 2018), and many other fields (for a review, see Wasserman, 2018). In cosmology, it has already been used to analyse the structure of the CMB (Pranav et al., 2017, 2019; Pranav, 2021) and the cosmic web (van de Weygaert et al., 2013; Pranav et al., 2017; Biagetti et al., 2020). Furthermore, an effective void finder has been built from the framework of persistent homology (Xu et al., 2019).

In this chapter, we want to establish persistent homology as an analysis tool that can be used to constrain cosmological parameters. In Sect. 4.2, we introduce the concept of persistent homology and describe how it can be used to constrain cosmological parameters. In particular, we focus on relating an intuitive understanding between the main output of persistent homology, the *persistence diagram*, to the underlying data. Afterwards, we compare it to other higher-order statistics that analyse cosmic shear data in a similar way, in particular peak count statistics and Minkowski functionals. That section is based on Heydenreich et al. (2021), which can be found in App. C.1.

In Sect. 4.3, we present our optimisation of the analysis setup, which includes different persistence statistics and various data compression methods. We then introduce a method to marginalise over systematic effects and conclude with a tomographic analysis of DES-Y1. That section is based on Heydenreich et al. (2022a), which can be found in App. C.2.

Throughout this chapter, we present a condensed summary of the published works Heydenreich et al. (2021) and Heydenreich et al. (2022a), glossing over some of the details and encountered problems. For a more complete picture, the interested reader is referred to the articles in App. C.1 and C.2.

For both sections, we work on signal-to-noise (S/N) maps of aperture mass that are computed from galaxy catalogues tracing the properties of a realistic survey. Before diving into the analysis using persistent homology, we want to briefly highlight how these maps were computed.

4.1. Calculating maps of aperture mass

Since both analyses in Sect. 4.2 and 4.3 utilise maps of aperture masses, we briefly explain how we compute aperture mass maps from a galaxy catalogue. Some analysis choices (such as the number of pixels of the aperture mass map) differ between Sect. 4.2 and 4.3, but the overall procedure is the same.

From a galaxy catalogue, we want to use Eq. (2.92) to calculate the aperture mass maps of a survey. To ease computation time, we distribute the galaxies on a pixel-grid[1]. In each pixel, we calculate the sum of all galaxy ellipticities ϵ_i and their squared moduli $|\epsilon_i|^2$. We then want to calculate the S/N maps of aperture masses as

$$\frac{S}{N}(\boldsymbol{\theta}) = \frac{\sqrt{2}\sum_i Q_{\theta_\mathrm{ap}}(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|)\epsilon_\mathrm{t}(\boldsymbol{\theta}_i; \boldsymbol{\theta} - \boldsymbol{\theta}_i)}{\sqrt{\sum_i |\epsilon_\mathrm{t}(\boldsymbol{\theta}_i; \boldsymbol{\theta} - \boldsymbol{\theta}_i)|^2 Q^2_{\theta_\mathrm{ap}}(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|)}}\,, \tag{4.1}$$

where the tangential ellipticity ϵ_t is defined in analogue to γ_t in Eq. (2.90). We can now calculate the numerator and denominator of this equation via a fast Fourier Transform (FFT), saving a significant amount of computation time. We want to focus on detecting peaks in the matter distribution, so we take a filter function that maximises the S/N of NFW-profiles (Schirmer et al., 2007):

$$Q_{\theta_\mathrm{ap}}(\theta) = \left[1 + \exp\left(6 - 150\frac{\theta}{\theta_\mathrm{ap}}\right) + \exp\left(-47 + 50\frac{\theta}{\theta_\mathrm{ap}}\right)\right]^{-1} \left(\frac{\theta}{x_\mathrm{c}\theta_\mathrm{ap}}\right)^{-1} \tanh\left(\frac{\theta}{x_\mathrm{c}\theta_\mathrm{ap}}\right)\,. \tag{4.2}$$

Here, we choose an aperture radius of $\theta_\mathrm{ap} = 12\farcm5$, which seemed to work well for Martinet et al. (2018). For each tile of the SLICS[2], there are empty regions where no galaxies are present, either due to the shape of the survey window or masked regions. We mask each pixel of the S/N map where the *effective area* within the aperture radius is less than 50% (meaning less than 50% of pixels within the aperture radius contain any galaxies) and assign it a value of $-\infty$. We ensure that each tile of the SLICS is surrounded by masked pixels so that neighbouring regions are treated as independent in the persistence calculations.

4.2. A feasibility study

This section is based on Heydenreich et al. (2021), which has been published in the journal Astronomy & Astrophysics, Volume 648, ID. A74. This work has been achieved in collaboration with Benjamin Brück and Joachim Harnois-Déraps. For this work, I led the data analysis; Benjamin Brück was in charge of the theoretical framework of persistent homology, and Joachim Harnois-Déraps assisted in the usage of the SLICS and cosmo-SLICS simulations. The published paper can be found in App. C.1.

[1] In Sect. 4.2, we use 1024^2 pixel, giving an area of $0\farcm6^2$ per pixel, in Sect. 4.3, we use 600^2 pixels with an area of $1'^2$.

[2] In Sect. 4.2, the SLICS mimic the KV450 survey; in Sect. 4.3, they mimic the DES-Y1 survey.

Before we use persistent homology to perform a cosmological parameter analysis, we need to establish that it is indeed sensitive to the underlying cosmology and that we can construct a suitable framework for the analysis. Within this section, we want to focus on three main topics of interest. The first is how to get a suitable data vector from persistent homology. The main output from persistent homology is a persistence diagram (compare Fig. 4.1). For a cosmological parameter analysis, however, we need a data vector of fixed dimension and a covariance for that data vector (as described in Sect. 2.3.3). The next problem is that we need a model of the data vector. Even for Gaussian random fields, such a model is virtually impossible to construct due to the highly non-linear nature of the persistence statistics (compare Pranav, 2021). Once these two problems are solved, weinvestigate how persistent homology compares to other higher-order statistics, both by intuitive understanding and quantitative analysis.

Throughout this section, we work on the SLICS and cosmo-SLICS that have been tailored to match the KV450 survey.

4.2.1. From the persistence diagram to a data vector

At first, we construct a persistence diagram for one realisation of the KV450-like SLICS. To do that, we take the S/N maps of all 17 tiles and concatenate these into one large map. Since each of the maps is surrounded by a mask, they are treated as independent in the persistence calculation, and the way we concatenate them does not matter. We then extract a persistence diagram from this aperture mass field, as described in Sect. 2.3.1[3]. To construct a data vector, we measure the *persistent Betti numbers* $\beta_n(t, t')$ of that persistence diagram. For a dimension n, the persistent Betti number $\beta_n(t, t')$ measures how many n-dimensional topological features[4] were born before the threshold t and are still alive at t'. In other words, the value of $\beta_n(t, t')$ is equal to the number of points in the persistence diagram which lie to the top-left of the point (t, t') (compare Fig. 4.1).

We now want to select a set of points (t, t') so that the data vector consisting of all $\beta_n(t, t')$ contains as much cosmological information as possible. In order to do that, we measure the persistent Betti numbers on a dense grid in all 26 cosmo-SLICS simulations.[5] As these simulations have been run with the same random seed for initial conditions, the difference between the cosmo-SLICS is mainly caused by their different cosmological parameters. We then measure the persistent Betti numbers on the same grid in all SLICS simulations. As these are run with the same cosmological parameters but different initial conditions, the difference between their Betti numbers is caused mainly by sample variance. The ratio of the difference between the cosmo-SLICS divided by the standard deviation of the SLICS is thus a good measure of the *cosmological information content* of the persistent Betti numbers at different points (t, t'). We inductively select the points with the highest cosmological information content in the following way: The first element of our data vector is the point with the highest cosmological information content. Afterwards, assuming we already have d evaluation points selected, let $\boldsymbol{x}_i$ be the corresponding d-dimensional data

[3] In practice, we use the python-library GUDHI.

[4] Each generator of a homology group is called a topological feature (compare Sect. 2.3.1.3).

[5] To ensure that we do not focus on numerical artifacts, we only sample points (t, t') where the mean of $\beta_n(t, t')$ over all SLICS is larger than 40.

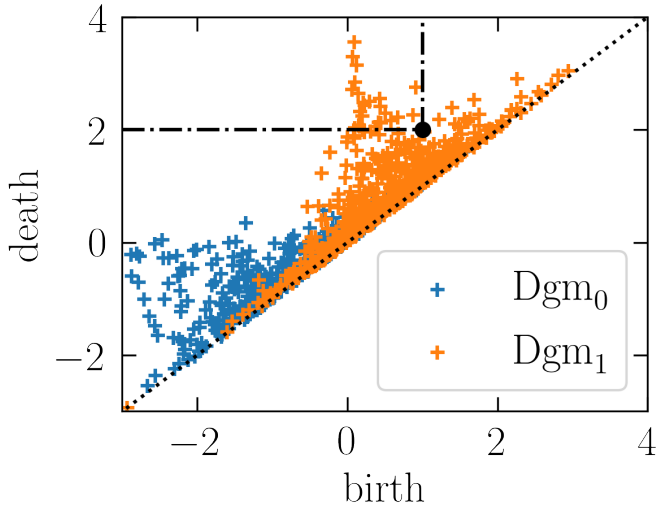


Figure 4.1. – An example persistence diagram of a S/N field of aperture masses. The blue dots signify the zero-dimensional features Dgm_0, corresponding to the connected components of the filtration. The orange dots correspond to the one-dimensional features Dgm_1, which are generated by the holes in the filtration of the S/N map. To evaluate the Betti function $\beta_1(1,2)$, one counts the number of 1-dimensional features to the top-left of the black point in the diagram.

vector in the cosmo-SLICS simulations, let $\boldsymbol{x}_{\mathrm{SLICS}}$ be the mean data vector of all SLICS-simulations, and let C be the covariance of $\boldsymbol{x}_{\mathrm{SLICS}}$. We then order the remaining evaluation points by their cosmological information content. Afterwards, for each point we construct new data vectors $\boldsymbol{x}'_i$, $\boldsymbol{x}'_{\mathrm{SLICS}}$ and a new covariance matrix C' that include the point with the next-highest cosmological information content. We then check if the inclusion of that point increases the quantity

$$\hat{\chi}^2 = \frac{1}{26} \sum_i \left[(\boldsymbol{x}'_{\mathrm{SLICS}} - \boldsymbol{x}'_i)^T C'^{-1} (\boldsymbol{x}'_{\mathrm{SLICS}} - \boldsymbol{x}'_i) \right] , \tag{4.3}$$

by at least 0.2.[6] If it does, the point gets accepted, and the procedure repeats with the new data vector that includes the new point. That way, we iteratively build our data vector until all remaining points are rejected.

Now we can generate a data vector of our summary statistic from an input galaxy catalogue. As a next step, we must construct a way to model this data vector for a set of cosmological parameters.

4.2.2. Constructing and validating a model for the data vector

As already stated, constructing a theoretical model for a statistic within the framework of persistent homology is not feasible. Instead, we rely on numerical simulations to achieve a model. For each set of cosmological parameters $\boldsymbol{\pi}_i$ in the cosmo-SLICS, we extract a data vector $\boldsymbol{x}_i$ and an error on the measurement $\boldsymbol{\sigma}(\boldsymbol{x}_i)$. We then build an emulator for our model via *Gaussian process regression (GPR)*. Briefly put, this GPR emulator

[6] The value of 0.2 was chosen more or less arbitrarily to ensure an appropriate length of the data vector.

provides a probabilistic way to interpolate the measured data vectors. For an arbitrary set of cosmological parameters $\boldsymbol{\pi}^\dagger$, the emulator provides both a model prediction $\boldsymbol{x}^\dagger$ and an uncertainty on that prediction, $\boldsymbol{\sigma}(\boldsymbol{x}^\dagger)$.

We then validate this emulator by a leave-one-out cross-validation: We remove the first of the 26 cosmologies of the cosmo-SLICS and train the emulator on the remaining 25 ones. Afterwards, we let the emulator predict the data vector $\boldsymbol{x}_1^\dagger$ for the first cosmology, which we had previously removed. We then calculate the error element-wise

$$\boldsymbol{\sigma}^2 = |\boldsymbol{x}_1^\dagger - \boldsymbol{x}_1|^2 \,, \tag{4.4}$$

where $\boldsymbol{x}_1$ is the measured data vector of the first cosmology. We verify that this error is smaller than the standard deviation we measure in the SLICS, and that it is consistent with the internal error estimate of the GPR emulator (compare App. C of Heydenreich et al., 2021).

This method has the advantage that it can predict our summary statistics with minimal computational effort and even provides us with an error estimate on the prediction. However, that error estimate can only capture the uncertainty of the prediction between different cosmologies of the cosmo-SLICS; if the simulations themselves are inaccurate, we have no way to quantify this.

It is important to note that the emulator only works in a highly specific setting: The persistent Betti numbers are not only sensitive to cosmological parameters and the redshift distribution of source galaxies but also to the exact shape of the survey footprint and both number density and shape noise of the source galaxies (compare App. D of Heydenreich et al., 2021). This means that the emulator and the N-body simulations used for training must be specifically adjusted to the survey properties.

4.2.3. Comparison to other higher-order statistics

We have established a way to perform a cosmological parameter analysis with persistent Betti numbers. However, the persistent Betti numbers are a relatively complicated statistic, and whether a cosmological parameter analysis using them is worth the effort is an important question. In this section, we thus want to compare persistent Betti numbers with peak count statistics (as used in Martinet et al., 2018); a relatively simple summary statistic.

We implement the peak count statistics on the same S/N maps of aperture masses by finding all local maxima in a map and binning them by their maximum value. To ensure compatibility with Martinet et al. (2018), we adapt their approach and use 12 bins between S/N values of 0 and 4.

4.2.3.1. Relation to peak count statistics

To compare persistent homology to peak count statistics, we first want to understand the physical interpretation of the topological features in the persistence diagram. To aid that understanding, we recommend the reader to re-visit Fig. 2.11 on page 56.

We start with the first homology group, which describes (1-dimensional) holes in the filtration. As seen in Fig. 4.2, a hole arises when the cut-off threshold is lower than the maximum value of an overdensity but higher than the value of its local environment. This

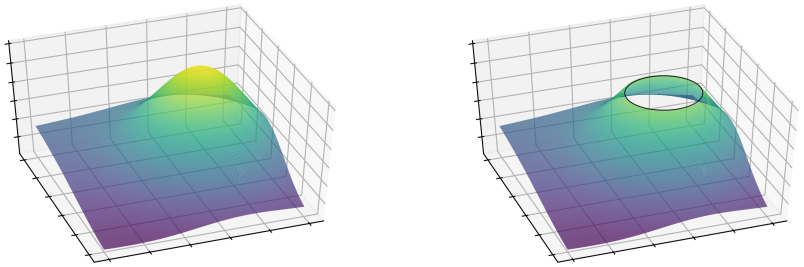


Figure 4.2. – A simplified example of a local overdensity in the S/N map (left), and the same overdensity when a cut-off threshold is applied (right). Adapted from Heydenreich et al. (2021).

0	0	0	0	0	0	0	0	0	0
0	0	4	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	2	0	0	0	0	0	0	3	0
0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	0
0	0	2	2	2	1	1	0	0	0
0	0	2	4	2	3	1	0	0	0
0	0	2	2	1	1	1	0	0	0
0	0	0	0	1	2	1	0	0	0
0	0	1	0	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0

Figure 4.3. – Two simplified S/N maps of aperture mass. Both maps contain four local maxima with the values 1, 2, 3, and 4. The maxima in the left map are completely isolated, whereas in the right map three of the maxima lie in a region with increased S/N. Adapted from Heydenreich et al. (2021).

hole disappears once the cut-off threshold exceeds the maximum value of said overdensity. This means that a hole corresponds to a local overdensity, where its birth relates to the S/N value of its local environment, and its death is equal to its maximum value. A similar line of reasoning shows that the 0th homology group, which describes connected components in the filtration, quantifies the behaviour of local minima: The birth of a connected component is equal to its minimum value, and its death (when it merges with other connected components) roughly corresponds to the S/N value of its environment. This already means that persistent homology contains at least as much cosmological information as peak statistics, as we can perfectly recover the latter by taking the death values of all features in the first homology group.

We argue that, aside from information about local minima, the information about the persistence of features gives an advantage to persistent homology over peak statistics. Inspecting Fig. 4.3, we can see that peak count statistics would detect four peaks of height 1, 2, 3, and 4 in both S/N maps. However, persistent homology would detect the following 1-dimensional features (denoted by their birth and death) in the left image: $\{(0,1),(0,2),(0,3),(0,4)\}$. In the right image, however, the detected features would be: $\{(0,1),(0,4),(1,2),(2,3)\}$. While these two maps would look indistinguishable under peak count statistics, persistent homology can detect a difference between them. Indeed, this information gain comes from the environment around the peaks: Peak count statistics

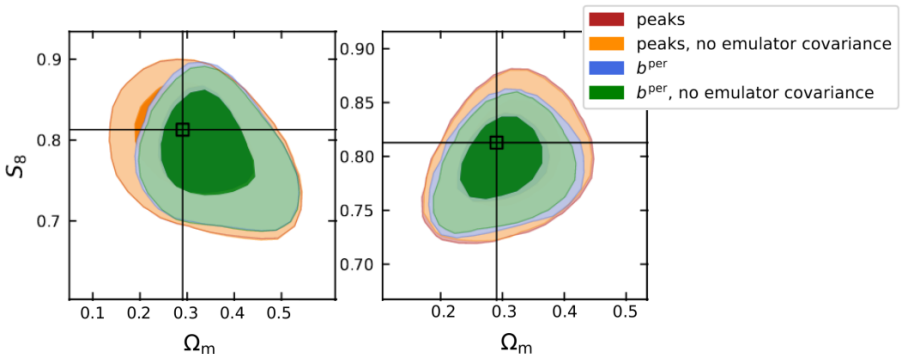


Figure 4.4. – Constraints on the cosmological parameters Ω_{m} and S_8 for a KV450-like setup (left) and a *Euclid*-like setup (right). In both cases, we compare the constraints from persistent Betti numbers (blue) and peak statistics (red). We also show the constraints that can be achieved when neglecting the emulator uncertainty (which could be achieved by significantly increasing the number of training nodes). We note that the effective area for the *Euclid*-like setup is only $100\,\mathrm{deg}^2$, so the absolute size of the contours should not be compared. Figure adapted from Heydenreich et al. (2021).

only utilise a peak's maximum value and thus cannot quantify how much a peak protrudes from its surroundings. In persistent homology, the birth of a feature yields exactly this information.

4.2.3.2. Comparison of cosmological parameter constraints

We conduct a cosmological parameter analysis on mock data to quantify the information gain from persistent homology. Our first analysis is performed for a realistic KV450-like setup that mimics the exact distribution of source galaxies and shape noise in the KV450 survey. To do that, we interpret the mean data vector measured in all SLICS, $\boldsymbol{x}_{\mathrm{SLICS}}$, as a measurement. Furthermore, we extract the sample covariance matrix C_{b} for $\boldsymbol{x}_{\mathrm{SLICS}}$ from the SLICS. At each step of the MCMC run, we take the diagonal matrix C_{e} whose entries correspond to the error estimates of the GPR emulator and define our covariance matrix C as $C = C_{\mathrm{b}} + C_{\mathrm{e}}$. We choose uniform priors for the cosmological parameters $\Omega_{\mathrm{m}}, S_8, h$ and w_0, and adapt the multivariate t-distribution from Sellentin and Heavens (2016), as discussed in Sect. 2.3.3.1.

We find that persistent homology yields a marginal increase of 3% in constraining power compared to peak statistics. However, the 'figure of merit' (i.e. one over the area covered by the 2σ contours, compare Albrecht et al., 2006), in the $\Omega_{\mathrm{m}} - \sigma_8$ plane increases by 48%. We also find that the emulator uncertainty plays a role and that its impact on persistent Betti numbers is worse than for peak count statistics.

We then repeat the analysis for a setup that mimics the properties of a Stage-IV survey. Our mock galaxy catalogues are constructed with a number density and source redshift distribution that are expected for the *Euclid* survey. In that setup we find an increase in constraining power of about 19% for S_8.

We do not compare the performance of persistent homology to other topological statistics like Minkowski functionals or the Euler characteristic. However, Zürcher et al. (2021) showed that these are already less constraining than peak count statistics. Since persistent

Betti numbers are more constraining than regular Betti numbers, we can conclude that persistent homology provides the most potent topological summary statistic for cosmic shear surveys to date.

4.3. Application to the Dark Energy Survey

This section is based on Heydenreich et al. (2022a), which was accepted[a] by the journal Astronomy & Astrophysics. This work has been achieved in collaboration with Benjamin Brück, Pierre Burger, Joachim Harnois-Déraps, Sandra Unruh, Tiago Castro, Klaus Dolag and Nicolas Martinet. I lead the data analysis, Benjamin Brück developed the necessary mathematical background, Pierre Burger was in charge of the measurement and validation of the shear 2pcf, Joachim Harnois-Déraps provided the suites of SLICS and cosmo-SLICS simulations, and Sandra Unruh developed the integration into CosmoSIS (Zuntz et al., 2015), which is a modular code to perform cosmological parameter analyses via MCMC. Tiago Castro, Klaus Dolag and Nicolas Martinet provided the post-processed Magneticum simulations, which were used to marginalise over baryonic effects. Of course, many ideas and analysis choices resulted from discussions with the collaborators, so it is impossible to disentangle the contributions perfectly. The accepted paper can be found in App. C.2.

[a] https://doi.org/10.1051/0004-6361/202243868

After showing in the feasibility study that persistent homology is a competitive tool for cosmological parameter analysis in cosmic shear surveys, we apply this method to the DES-Y1 survey. We improve upon the previous study in several ways. Our first improvement is to perform a tomographic analysis. Furthermore, we abandon the previously used persistent Betti numbers in favour of a more stable summary statistic: persistent heatmaps. Afterwards, we test different data compression methods on the heatmaps and improve upon the previously developed method to choose evaluation points. Lastly, we implement marginalisation strategies for astrophysical and observational systematics (see Sect. 2.2.6.2). These improvements allow us to perform a cosmological parameter analysis of the DES-Y1 survey, yielding the first parameter constraints ever achieved within the framework of persistent homology.

4.3.1. A tomographic analysis

The galaxies in the DES-Y1 survey are split into four tomographic redshift bins based on their photometric redshifts. A tomographic analysis thus allows us to trace the growth of structure over cosmic time. The 'naive' approach to such a parameter analysis would be calculating the aperture mass maps for all four tomographic bins and performing the persistent homology calculations on those four maps. However, important information is missed when utilising this approach. For example, assuming we have a nearby massive galaxy cluster, this would show up as a local maximum in all four aperture mass maps. However, the information that this local maximum is at the same position in all four maps

would get lost as soon as one calculates the persistence diagram of these maps. Instead of just taking all four tomographic bins, we thus also calculate the aperture mass maps of all *combinations* of tomographic bins $(1, 2, 3, 4, 1\cup2, \ldots, 3\cup4, \ldots, 1\cup2\cup3\cup4)$. Martinet et al. (2021) showed that this method has the potential to almost double the constraining power on the structure growth parameter S_8, which has been confirmed by Harnois-Déraps et al. (2021, hereafter HD+21). However, taking all combinations of the four tomographic bins yields $2^4 - 1 = 15$ potential combinations, which restricts the possible length of a data vector for one tomographic bin combination.

4.3.2. Optimising the persistence statistics

As the next step, we investigate different persistence statistics. In the previous study, we used persistent Betti numbers, some of the most straightforward statistics in persistent homology. However, they come with some disadvantages. In particular, the value of Betti functions is always an integer, which means that usually, the average of several Betti functions is not a Betti function anymore. Instead, we opt to use *persistent heatmaps*. To construct these, we start with the set of all features $\{(b_i, d_i)\}_i$ in the persistence diagram. We then represent each feature by a Dirac delta distribution $\delta_\mathrm{D}\left[\boldsymbol{x} - (b_i, d_i)^T\right]$ and define the persistence heatmap as the convolution of these Dirac delta distributions with an isotropic Gaussian with standard deviation t. To ensure that the heatmap is zero on the diagonal, for each feature (b_i, d_i) we add a negative Dirac delta distribution $-\delta_\mathrm{D}\left[\boldsymbol{x} - (d_i, b_i)^T\right]$ at the point (d_i, b_i) that is mirrored along the diagonal. This is motivated by the fact that features close to the diagonal are likely to be caused by tiny noise fluctuations.

These persistence heatmaps can be understood as a smoothed persistence diagram that is obtained by just replacing every feature with a Gaussian. The smoothing parameter t is an important hyperparameter that needs to be chosen carefully; a too small value of t might lead to unstable results, whereas a too large value likely results in loss of constraining power. For our purposes, we find that $t = 0.2$ yields a good compromise.

We note that each persistence diagram Dgm results in two heatmaps, one for the 0-dimensional features Dgm_0, and one for the 1-dimensional features Dgm_1.

We also try to use persistence landscapes (Bubenik, 2015), a more elaborate statistic in persistent homology. However, we are unable to set them up in a way that leads to stable and competitive results.

4.3.3. Optimising the data compression method

With the persistent heatmaps as our chosen summary statistic, we still needed to choose a data compression method. We opt for an optimisation of the method introduced in Sect. 4.2.1. For each combination of tomographic bins, we sample each heatmap at 100×100 points and take the upper diagonal matrix of this regular grid (see Fig. 4.5). We then calculate the cosmological information content of each point, as described above. Again, we inductively build our data vector and start by selecting the point with the highest cosmological information content. Given a data vector $\boldsymbol{x}$, for each point in the heatmap, we build a data vector $\boldsymbol{x}'$ by adding that point to the existing one. We then extract that data

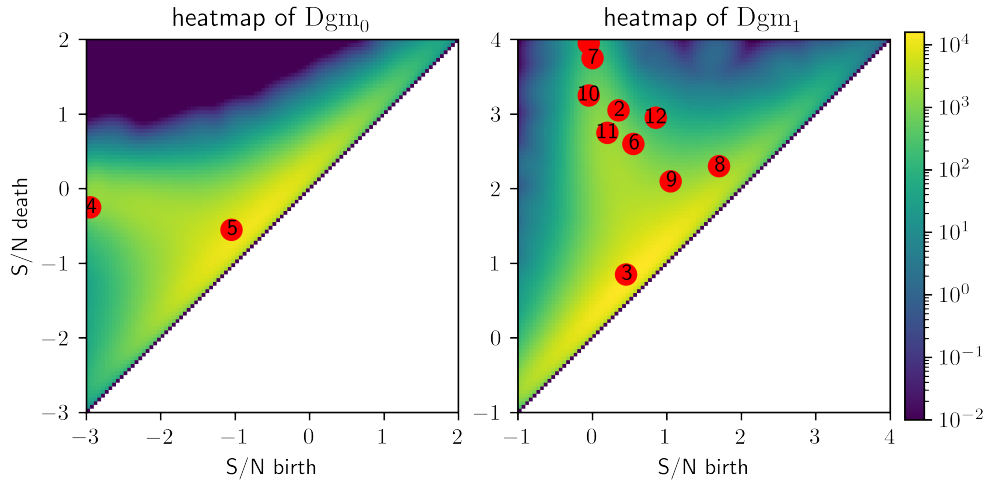


Figure 4.5. – An example of a heatmap for the combination of all four tomographic bins in one line of sight of the DES-Y1-like SLICS. On the left side, we show the heatmap for Dgm$_0$; the right panel shows the heatmap for Dgm$_1$. The red dots represent the evaluation points chosen by our data compression method. From Heydenreich et al. (2022a).

vector from all cosmo-SLICS simulations and define those as $\boldsymbol{x}'_{\mathrm{cosmoSLICS},i}$. Furthermore, we extract the mean of the cosmo-SLICS, $\left\langle \boldsymbol{x}'_{\mathrm{cosmoSLICS},i} \right\rangle_i$ and define $\Delta\boldsymbol{x}'_i = \boldsymbol{x}'_{\mathrm{cosmoSLICS},i} - \left\langle \boldsymbol{x}'_{\mathrm{cosmoSLICS},i} \right\rangle_i$. We then also extract the covariance of the data vector $\boldsymbol{x}'$, C', from the SLICS. Afterwards, we choose the point for which the new data vector maximises

$$\hat{\chi}^2 = \sum_{i=0}^{25} \Delta\boldsymbol{x}_i'^T C'^{-1} \Delta\boldsymbol{x}'_i \,. \tag{4.5}$$

Instead of setting an arbitrary threshold by which $\hat{\chi}^2$ needs to increase (as in Sect. 4.2.1), this method ensures that it takes the best available point to maximise the cosmological information content of a data vector. We decide to take 12 points of the heatmap per tomographic redshift bin, resulting in a data vector of length 180 and ensuring compatibility with HD+21.

We compare this data compression method with two others: a principal component analysis (PCA) and a method developed by Asgari and Schneider (2015). The former has the advantage that it can be used on data vectors of any length and identifies linear combinations of elements that cause the majority of variation between data vectors. However, it has the significant disadvantage that it can not take into account the sample covariance of the data vector. In other words, it maximises the variation of a data vector, not its cosmological information content. Conversely, the method from Asgari and Schneider (2015) tries to maximise the cosmological information content in a framework extending the Fisher formalism; however, this requires knowledge of the inverse covariance matrix, which can not be obtained for arbitrarily long data vectors (Hartlap et al., 2007). Both data compression methods do not manage to out-perform the method we developed but instead yield results that are very consistent with our fiducial ones.

4.3.4. Marginalising over systematic effects

While we have already constructed a modelling pipeline for persistent homology statistics with a GPR emulator, which can be easily adapted to the specifications of the DES-Y1 survey, we have not yet taken into account the systematic biases that are inherent to every cosmic shear survey. We use specifically tailored N-body simulations to model the effects of intrinsic alignments, baryon feedback, multiplicative shear measurement bias and photometric redshift errors.

For the intrinsic alignments, we use SLICS simulations where galaxies are placed at the centres of dark matter halos and then infused with the non-linear intrinsic alignment model with the intrinsic alignment amplitudes $A \in \{-5, -2, -1, -0.5, 0.5, 1, 2, 5\}$. For the baryonic feedback, we utilise the magneticum hydrodynamical simulations, which have been run on the same initial conditions both with and without baryonic feedback processes, allowing us to determine the impact of this effect on the persistent heatmaps. To estimate the impact of multiplicative shear measurement biases, we multiply the simulated shear in the cosmoSLICS fiducial cosmology by a factor of $(1 + m)$, where $m \in \{-0.025, -0.0125, 0.0125, 0.025\}$ parametrises the measurement bias. Lastly, to account for photometric redshift uncertainties, we create ten 'biased' source redshift distributions for each tomographic bin, where we shift the fiducial redshift distribution by a random number drawn from a normal distribution whose variance corresponds to the estimated photometric redshift uncertainty.

For every systematic bias, we assume that the effect is linear and does not depend on the underlying cosmology. This is certainly a simplified assumption, but one that we believe to be reasonably accurate. For every point in the data vector x and every systematic bias, we can now perform a linear fit

$$\boldsymbol{x}_{\mathrm{sys}}(\lambda) = \boldsymbol{m}_{\boldsymbol{x}}\lambda + \boldsymbol{x}_{\mathrm{nosys}} \ , \tag{4.6}$$

where $\boldsymbol{x}_{\mathrm{nosys}}$ is the fiducial data vector, $\boldsymbol{x}_{\mathrm{sys}}$ is the one that is contaminated by systematic effects, and λ parametrises the strength of the respective systematic effect. We now further assume that the effects of each systematic bias are independent of each other and model our full systematics-infested data vector as

$$\boldsymbol{x}_{\mathrm{sys}} = \boldsymbol{x}_{\mathrm{nosys}} + \boldsymbol{m}_{\mathrm{IA}}A_{\mathrm{IA}} + \boldsymbol{m}_{\mathrm{bar}}b_{\mathrm{bar}} + \boldsymbol{m}_{\mathrm{dz}}\Delta z + \boldsymbol{m}_{\mathrm{m}}\Delta m \ , \tag{4.7}$$

where $A_{\mathrm{IA}}, b_{\mathrm{bar}}, \Delta z$ and Δm parametrise the effects of intrinsic alignments, baryons, redshift errors and shear measurement biases, respectively.

We test the impact of the individual systematic effects in two ways: First, we run a cosmological parameter analysis on a systematics-infused data vector without marginalising over any systematics to assess the bias this effect would cause when left untreated. Afterwards, we run a parameter analysis where we marginalise over this systematic effect and compare it to the un-marginalised case to quantify the loss in constraining power we suffer by including this systematic effect. For intrinsic alignments and baryonic feedback, these impacts can be seen in Fig. 4.6; for shear measurement errors and photometric redshift uncertainties, the impact is not significant. We can see that both baryonic feedback and intrinsic alignments significantly impact our parameter constraints and must be treated carefully. We also note that marginalisation over both systematic effects significantly increases the uncertainty in the cosmological parameter constraints.

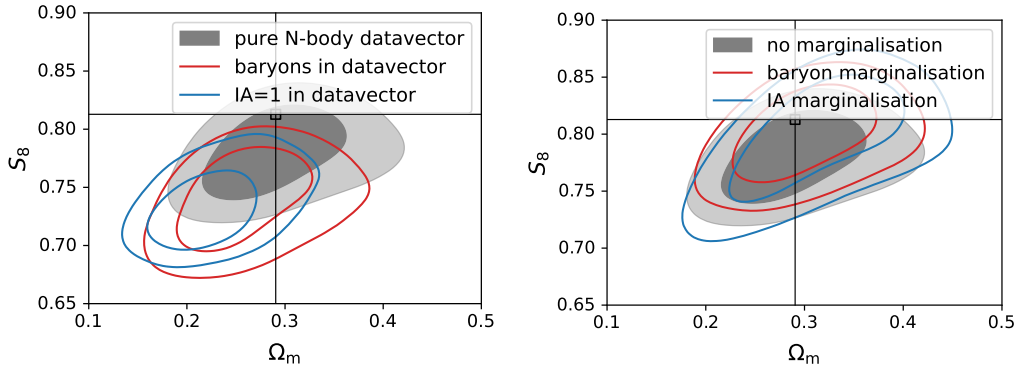


Figure 4.6. – The left figure depicts the impact of systematic biases on cosmological parameters if they are present in the target data vector and not properly accounted for. The right figure shows the loss in constraining power when one marginalises over one of the systematic effects. Adapted from Heydenreich et al. (2022a)

4.3.5. Results

Aside from an analysis with persistent homology, we also include an analysis with the shear 2pcf $\xi_{\pm}$, replicating the analysis settings chosen in Troxel et al. (2018). We measure the 2pcf in the SLICS to extract a joint covariance matrix between persistent homology and shear 2pcf and integrate our GPR emulator into the public CosmoSIS pipeline. This allows us to jointly model both shear 2pcf and persistent homology, where the former are modelled from a theoretical power spectrum (Takahashi et al., 2012) and the latter from the GPR emulator. For the 2pcf, the scale-cuts were chosen to ensure minimal containment from baryonic feedback; all other nuisance parameters can be accounted for within the CosmoSIS pipeline. To ensure both a reasonable runtime and reliable parameter constraints, we choose the nested sampling method polychord (Handley et al., 2015) with the settings outlined by Lemos et al. (2022).

Before conducting a cosmological parameter analysis, we validate our analysis pipeline on mock data to verify that we recover the input cosmology. We take the mean data vector measured in the SLICS as a target data vector and perform a complete parameter analysis, including all systematic effects.

The results of the validation tests can be seen in Fig. 4.7. We see that we recover the input cosmology well within our 1σ uncertainties. We can further see that the constraints from persistent homology appear to be tighter than the ones from shear 2pcf and that a joint analysis further increases the constraining power. We conclude that our validation test has been successful and that our analysis pipeline is ready to be applied to real data.

We show the parameter constraints achieved from the DES-Y1 survey in Fig. 4.8. We see that persistent homology and shear 2pcf yield perfectly consistent constraints for the matter clustering parameter S_8 and the intrinsic alignment amplitude A, although the constraints achieved by persistent homology are significantly tighter. Neither of the two methods appears to place any significant constraints on the dark energy EoS w_0. We notice a large tension between shear 2pcf and persistent homology in the constraints of the matter density parameter Ω_m, which unfortunately prevents us from performing a joint analysis. We see that the constraints from HD+21, achieved with an independent

analysis pipeline using peak count statistics, are remarkably similar to ours.[7] This leads us to believe that this tension is not caused by a bug in our analysis setup. We tested different potential causes of this bias: We try excluding specific nodes of our cosmoSLICS training set, applying different scale cuts to the persistence heatmaps, or omitting certain tomographic bins. To check whether the underlying cosmoSLICS training set of the GPR emulator might be responsible for this shift, we also try modelling the shear 2pcf using the same GPR emulator. None of these tests reduce the Ω_m tension by a significant amount.

[7] We note that, in contrast to the peak count statistics analysis of HD+21, we marginalise over baryonic effects and intrinsic alignments, which increases our posterior contours (compare Fig. 4.6). This explains why we do not achieve tighter constraints despite using a more powerful summary statistic.

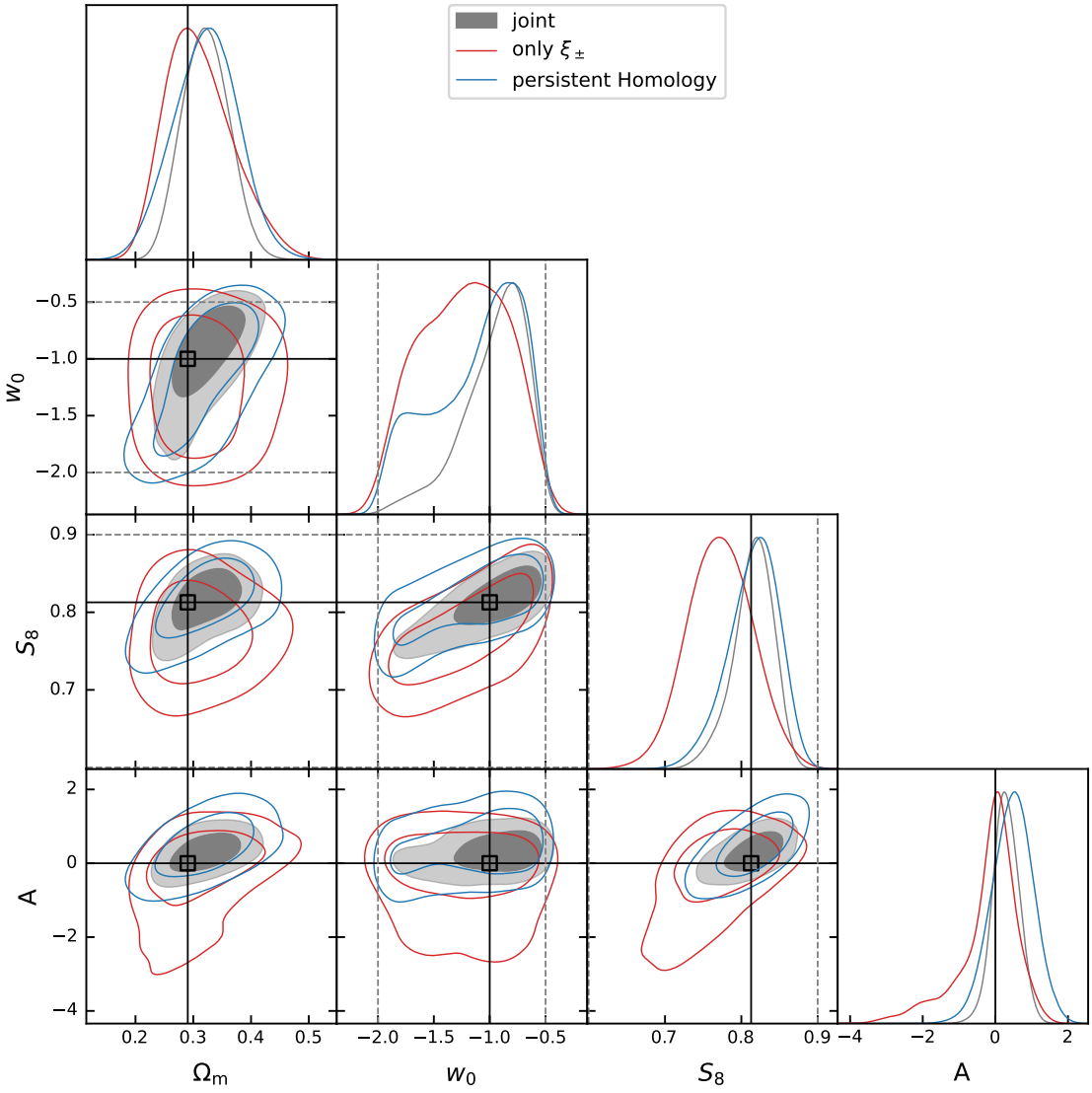


Figure 4.7. – Parameter constraints of an analysis of a DES-Y1-like mock data vector with known cosmological parameters. We show the constraints from shear 2pcf (red) and the ones from persistent homology (blue). We further show the constraints of a joint analysis (grey). The black crosses depict the cosmological parameters of the cosmoSLICS training nodes; the solid lines show the cosmological parameters of the mock data vector. The dashed lines represent the boundaries imposed by the priors. From Heydenreich et al. (2022a).

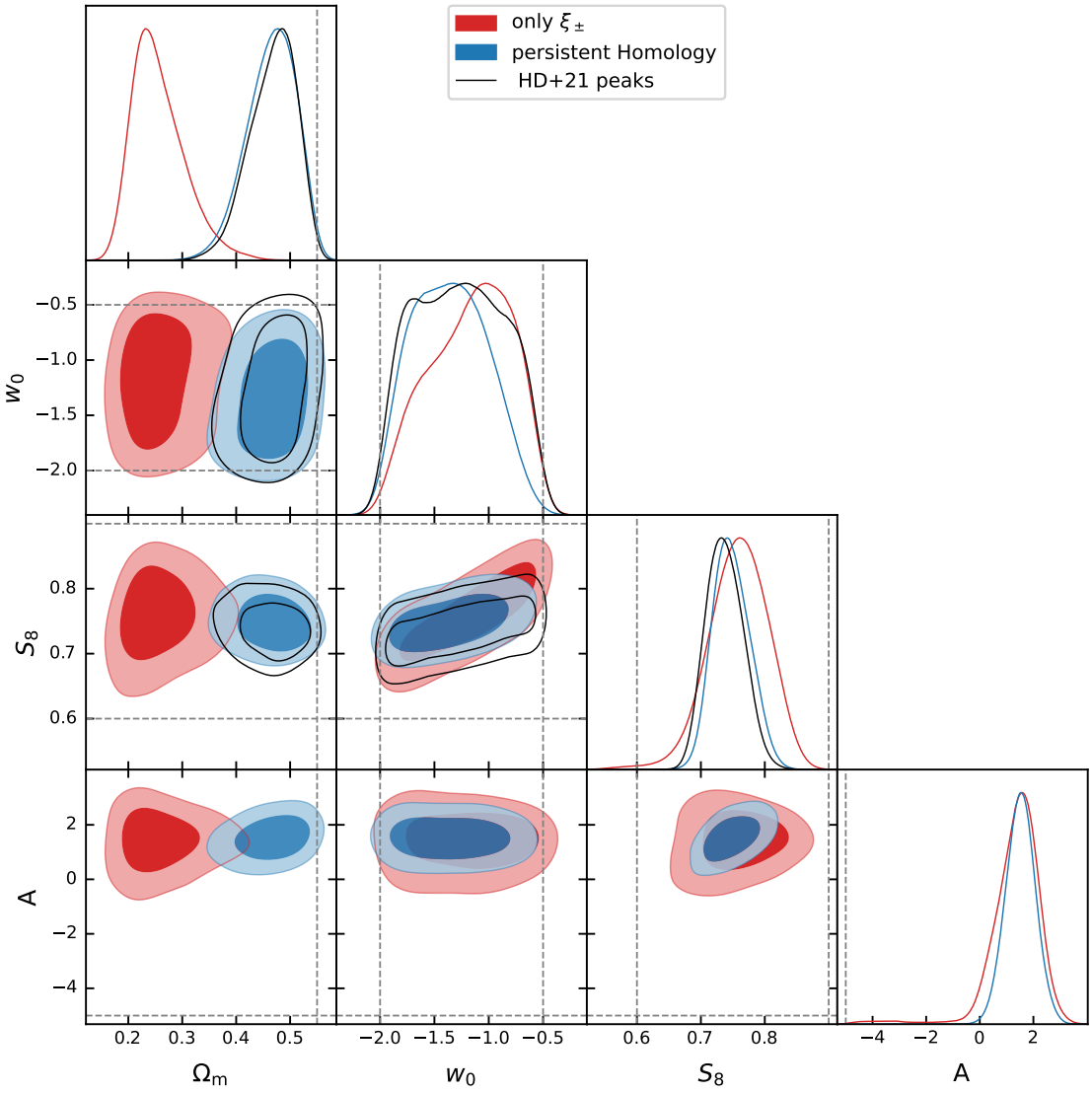


Figure 4.8. – Parameter constraints of an analysis of the DES-Y1 survey. We show the constraints from shear 2pcf (red) and the ones from persistent homology (blue). The black crosses depict the cosmological parameters of the cosmoSLICS training nodes; the dashed lines represent the boundaries imposed by the priors. For comparison, we also show the constraints from peak statistics that HD+21 achieved with an independent analysis setup (black). From Heydenreich et al. (2022a).

4.4. Concluding remarks

In this chapter, we have introduced a way to perform a cosmological parameter analysis with persistent homology. We show that this analysis can be performed on real survey data and that it out-performs other commonly used higher-order statistics, such as peak count statistics, both by an intuitive understanding and a quantitative analysis. The analysis setup includes many choices (such as the filter radius for the aperture mass maps θ_{ap}, the smoothing scale of the heatmap, the number of evaluation points, ...), and with careful fine-tuning, one could probably achieve even better results.

Our constraints for the matter clustering parameter S_8 and the intrinsic alignment amplitude A are consistent with the ones from second-order statistics, albeit significantly tighter. In particular, we can report a detection of the intrinsic alignment effect at almost 3σ. In the matter density parameter Ω_{m}, however, we measure a significant tension between 2pcf and persistent homology, which is unlikely to be caused by the analysis setup itself. The origin of this tension is not yet understood. It might just be an unlikely statistical fluctuation in the DES-Y1 data or the manifestation of some untreated systematic effect. Curiously, something similar has happened in other analyses: When investigating HSC SSP data with both real- and Fourier-space statistics, Hamana et al. (2020) find a remarkable consistency in the parameter S_8 and a tension in Ω_{m} (albeit not as large as ours by far). However, a more exciting explanation might be that persistent homology (or any other field-level higher-order statistic) is sensitive to some non-Gaussian properties of the LSS that two-point statistics can not capture. For example Biagetti et al. (2022) found that persistent homology is very sensitive to non-Gaussianities in the primordial seeds of structure formation.

Curiously, the results from the DES-Y1 highlight both the advantages and disadvantages of an analysis performed with persistent homology. In raw constraining power, the method can out-perform the commonly used two-point statistics while appearing to remain unbiased. However, once we obtain results we do not expect, it is hard to test the summary statistics rigorously. Since the summary statistics explicitly depend on the survey footprint and galaxy number density, comparing them between different simulation setups is difficult. Furthermore, the GPR emulator resembles a 'black box' that outputs the desired summary statistics without allowing us to gain a meaningful physical understanding. If artefacts in the simulations contaminate the statistics, we cannot find out. In other words, parameter constraints achieved by persistent homology (or any simulation-based inference) are only as good as our confidence in these constraints, and our options to gain that confidence are limited.

For future work, we plan to perform a similar analysis of the KiDS-1000 survey. If we do not find a tension in Ω_{m} there, we can likely attribute this tension to a statistical fluctuation in the DES-Y1 data. If we measure similar results for the matter density parameter, we will have to investigate the reason behind this discrepancy thoroughly. One test that should be performed in that case is to build an emulator based on a different simulation suite, like the BACCO simulation project (Angulo et al., 2021) and see if it is consistent with the one based on the cosmoSLICS.

CHAPTER 5

Third-order statistics

This chapter discusses our efforts to prepare a cosmological parameter analysis with third-order shear statistics. The potential information gain of a joint analysis has already been forecasted (Kilbinger and Schneider, 2005; Kayo et al., 2013; Sato and Nishimichi, 2013) and a first cosmological parameter analysis has been performed (Fu et al., 2014). However, as established in the previous chapter, achieving tight constraints on cosmological parameters is only useful if one can be confident in those constraints and the methods used to achieve them. Especially in light of the emerging S_8 tension, high confidence in cosmic shear analyses is more important than ever. In this work, we thus want to focus on rigorously testing and validating third-order cosmic shear statistics.

The first section of this chapter describes our efforts to model the shear 3pcf $\Gamma^{(i)}$ and third-order aperture statistics $\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle$. We outline the modelling pipeline we built and the tests we performed to ensure reliability. This section is based on the as of yet not peer-reviewed article Heydenreich et al. (2022b).

The second section of this chapter focuses on developing a model for the covariance of third-order aperture statistics. Laila Linke led this effort, however, it built upon my work, and I made significant contributions to the project. We plan to publish the content of the second section soon (Linke et al. in prep.).

We also include a measurement of the shear 3pcf and third-order aperture masses for KiDS-1000 in App. B, which we have not included in the paper since we are so far unable to account for the astrophysical systematics that bias this measurement.

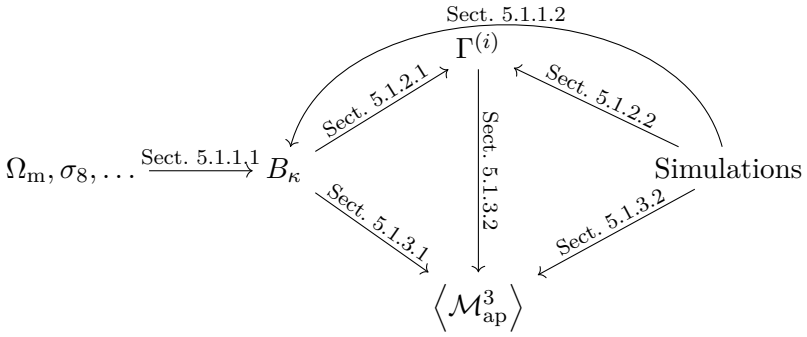


Figure 5.1. – A schematic diagram of the modelling and validation pipeline introduced in this section. The numbers on the arrows correspond to the respective section where this part of the pipeline is discussed. Adapted from Heydenreich et al. (2022b).

5.1. Modelling and validation

This section is based on the not-yet reviewed article Heydenreich et al. (2022b), which has been submitted to the journal Astronomy & Astrophysics. This work has been conducted in collaboration with Laila Linke, Pierre Burger and Peter Schneider. Here, I implemented the modelling algorithms for BiHalofit and the shear 3pcf as well as the interface to treecorr. Furthermore, I developed the algorithms to measure the bispectrum and the third-order aperture statistics, as well as most of the performed tests, including the generation of Gaussian and lognormal random fields. Laila Linke implemented the modelling algorithm for third-order aperture statistics and made countless other significant contributions to the modelling pipeline, including but not limited to creating the interface to the cosmoSIS pipeline. Pierre Burger was responsible for the mock-MCMC runs, including the training of the CosmoPower emulator and everything that relates to the T17 simulations we used in this work. Peter Schneider provided invaluable insight into many aspects of a cosmological analysis with third-order statistics. Of course, most of the statements in this section arise from countless discussions with all collaborators, making it impossible to disentangle the contributions made to this work perfectly.

In addition to the efforts outlined in this section, I found and fixed a bug in the BiHalofit routine, which would yield incorrect results if the three k-input values were not in ascending order. Furthermore, our tests revealed a bug in the treecorr algorithm, which would yield wrong results for the conversion from shear 3pcf to third-order aperture mass statistics of non-equal scale, which Laila Linke managed to identify and fix[a]. Lastly, we performed additional validation tests, in particular for the treecorr algorithm, which we outline in App. A, where we further identified a bug in treecorr that would cause it to run incorrectly if the input galaxy catalogue was taken from a regular grid.

[a] Fixed in version 4.2.3, see the changelog for treecorr.

The main effort of this work was to develop and test the modelling pipeline for third-order shear statistics, which is visualised in Fig. 5.1. This pipeline is publicly available at `https://github.com/sheydenreich/threepoint/releases/`.

5.1.1. Convergence bispectrum

5.1.1.1. Modelling

To model the matter bispectrum (Eq. 2.51), we use the BiHalofit algorithm (Takahashi et al., 2020). In particular, for squeezed triangle configurations with $\ell_1 \approx \ell_2 \gg \ell_3$, this algorithm substantially improves over previous formulae such as in Scoccimarro and Couchman (2001); Gil-Marín et al. (2012). Where the latter two are subject to errors of over 200%, BiHalofit retains an accuracy of 10-20% or better. As can be seen in Halder et al. (2021), the effect on higher-order shear statistics can be substantial, especially on small scales $\lesssim 30'$. We then use the Limber-integration (Eq. 2.119) to calculate a model for the convergence bispectrum B_κ. The advantage of using BiHalofit is not only that we get a bispectrum model which is accurate to 10-20% down to small scales but also that the non-linear scales that we are required to compute for this bispectrum model are the same as the ones of the revised halofit model (Takahashi et al., 2012). This fact can save computation time in a joint modelling of second- and third-order shear statistics. Furthermore, the bispectrum is calculated from the linear power spectrum, which is known to have a very high degree of accuracy; Scoccimarro and Couchman (2001); Gil-Marín et al. (2012) use the non-linear power spectrum as an input, which can vary up to a few per cent between different models. To model the linear power spectrum, we mostly use the fitting formula from Eisenstein and Hu (1999); we compare the resulting statistics by using a power spectrum from CAMB as input and see that the difference is negligible.

5.1.1.2. Measuring

Although not strictly necessary, we want to validate the Limber-integrated BiHalofit bispectrum before integrating it to yield the other third-order statistics. To measure the bispectrum in N-body simulations, we adapt an estimator developed for three-dimensional matter fields by Watkinson et al. (2017) to two-dimensional convergence fields.

Given a convergence field $\kappa(\boldsymbol{\theta})$ and its Fourier transform $\tilde{\kappa}(\boldsymbol{\ell})$, for an ℓ-bin $\bar{\ell}_i = [\ell_{\min}, \ell_{\max}]$ we define $\tilde{\kappa}(\boldsymbol{\ell}; \bar{\ell}_i)$ as

$$\tilde{\kappa}(\boldsymbol{\ell}; \bar{\ell}_i) = \begin{cases} \tilde{\kappa}(\boldsymbol{\ell}) & \ell_{\min} \leq |\boldsymbol{\ell}| < \ell_{\max} \\ 0 & \text{otherwise} \end{cases} , \tag{5.1}$$

and $\kappa(\boldsymbol{\theta}; \bar{\ell}_i)$ as its inverse Fourier transform. We also define $I(\boldsymbol{\theta}; \bar{\ell}_i)$ as the inverse Fourier transform of $\tilde{I}(\boldsymbol{\ell}; \bar{\ell}_i)$ with $\tilde{I}$ defined as

$$\tilde{I}(\boldsymbol{\ell}; \bar{\ell}_i) = \begin{cases} 1 & \ell_{\min} \leq |\boldsymbol{\ell}| < \ell_{\max} \\ 0 & \text{otherwise} \end{cases} . \tag{5.2}$$

The estimator for the convergence bispectrum is then defined as

$$\widehat{B_\kappa}(\bar{\ell}_1, \bar{\ell}_2, \bar{\ell}_3) = \frac{\Omega^2}{N_{\text{pix}}^3} \frac{\sum_{i=1}^{N_{\text{pix}}} \kappa(\boldsymbol{\vartheta}_i; \bar{\ell}_1)\kappa(\boldsymbol{\vartheta}_i; \bar{\ell}_2)\kappa(\boldsymbol{\vartheta}_i; \bar{\ell}_3)}{\sum_{i=1}^{N_{\text{pix}}} I(\boldsymbol{\vartheta}_i; \bar{\ell}_1)I(\boldsymbol{\vartheta}_i; \bar{\ell}_2)I(\boldsymbol{\vartheta}_i; \bar{\ell}_3)} , \tag{5.3}$$

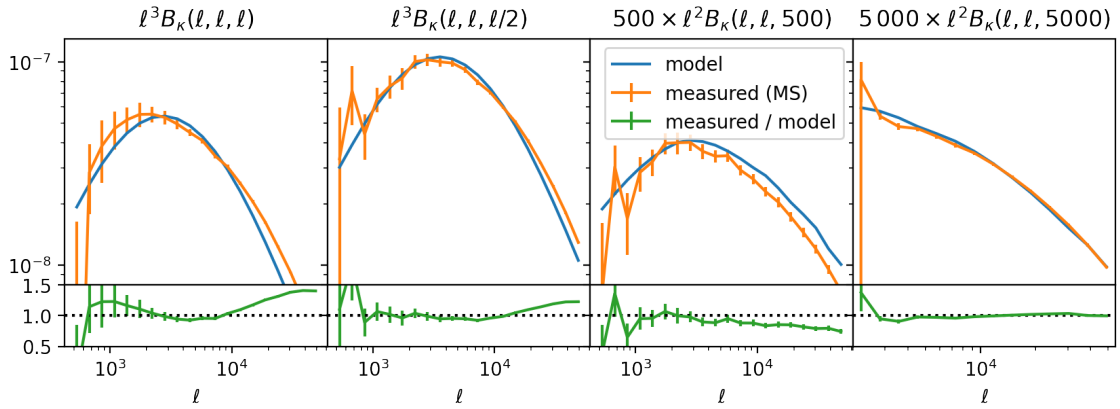


Figure 5.2. – In the top panels, we compare the model predictions of the convergence bispectrum (blue) to its measurements in the MS (orange) at redshift $z = 1$. In the bottom panels, we show the ratio of measurement and model prediction. The error bars signify the error on the mean of the measurements. We measure the bispectrum in a bin size of $\Delta\ell = 0.13\ell$ and average our model predictions over the same bin. In the latter case, we use equation (8) of Joachimi et al. (2009) to weigh the different triangle combinations within one bin.

where Ω is the solid angle of the respective field and $N_{\rm pix}$ is the number of pixels. Since the $I(\boldsymbol{\theta}; \bar{\ell}_i)$ can be pre-computed, and the $\kappa(\boldsymbol{\theta}; \bar{\ell}_i)$ only need to be computed once per bin $\bar{\ell}_i$, a measurement of the full bispectrum of a convergence field can be performed with only a few Fourier transformations. The limits of this estimator are that it assumes periodic boundary conditions, which are usually not fulfilled in ray-tracing simulations. Thus, when approaching either field-size or pixel-size scales, one should not trust the results of this estimator. In our approach, we discard scales smaller than 5 pixels or larger than a third of the field size.

5.1.1.3. Validation

We test the modelled convergence bispectrum with the MS. For each of the 64 lines of sight, we measure the convergence bispectrum for sources at $z \approx 1$ with the abovementioned estimator. We then compare those measurements to our model predictions in Fig. 5.2 and see that the convergence bispectrum is accurate within the sample variance over all tested triangle configurations for scales down to $\ell \lesssim 10^4$. For larger wavenumbers ℓ, we notice up to 40% discrepancy. While this might point to an inaccuracy in the BiHalofit model, it may very well be caused by the limited sample size of the MS[1] or the smoothing induced by the ray-tracing method (Hilbert et al., 2009).

[1]Since the 64 lines of sight were taken from a single simulation box, the sample variance between the lines of sight probably underestimates the true sample variance.

5.1.2. Shear three-point correlation functionss

5.1.2.1. Modelling

To model the shear 3pcf from the convergence bispectrum, we utilise Eqs. (2.130) and (2.131). In both cases, the integrals are complicated to solve numerically since they filter the bispectrum with the strongly oscillating 6-th or 2-nd order Bessel functions $J_{6/2}$. To mitigate this issue, we introduce polar coordinates $R = \sqrt{\ell_1^2 + \ell_2^2}$, $\psi = \arctan(\ell_2/\ell_1)$ and isolate the integration over the Bessel functions. The quantities in Eq. (2.132) transform to

$$\begin{aligned}
A_3' &= \frac{A_3}{R} = \sqrt{(x_2\cos\psi)^2 + (x_1\sin\psi)^2 + x_1x_2\sin 2\psi\,\cos(\varphi+\phi_3)}\,,\\
\cos 2\bar{\beta} &= (\cos\varphi + 2\cos\psi\,\sin\psi)\,,\\
\sin 2\bar{\beta} &= (\cos^2\psi - \sin^2\psi)\sin\varphi\,,\\
A_3'\cos\alpha_3 &= (x_2\cos\psi + x_1\sin\psi)\cos\left(\frac{\varphi+\phi_3}{2}\right)\,,\\
A_3'\sin\alpha_3 &= (x_2\cos\psi - x_1\sin\psi)\sin\left(\frac{\varphi+\phi_3}{2}\right)\,.
\end{aligned} \tag{5.4}$$

and the integrals can be written as

$$\begin{aligned}
\Gamma^{(0)}(x_1,x_2,x_3) = \frac{1}{3(2\pi)^3}\int_0^{2\pi}\mathrm{d}\varphi\int_0^{\frac{\pi}{2}}\mathrm{d}\psi\int_0^{\infty}\mathrm{d}R\;\; R^3\sin(\psi)\cos(\psi)\,b(R\cos(\psi),R\sin(\psi),\varphi)\,\mathrm{e}^{2\mathrm{i}\bar{\beta}}\\
\times\left[\mathrm{e}^{\mathrm{i}(\phi_1-\phi_2-6\alpha_3)}J_6(R\,A_3') + \mathrm{e}^{\mathrm{i}(\phi_3-\phi_2-6\alpha_1)}J_6(R\,A_1') + \mathrm{e}^{i(\phi_3-\phi_1-6\alpha_2)}J_6(R\,A_2')\right]\,,
\end{aligned} \tag{5.5}$$

$$\begin{aligned}
\Gamma^{(1)}(x_1,x_2,x_3) = \frac{1}{3(2\pi)^3}\int_0^{2\pi}\mathrm{d}\varphi\int_0^{\frac{\pi}{2}}\mathrm{d}\psi\int_0^{\infty}\mathrm{d}R\;\; R^3\sin(\psi)\cos(\psi)\,b(R\cos(\psi),R\sin(\psi),\varphi)\\
\times\Big[\mathrm{e}^{\mathrm{i}(\phi_1-\phi_2+2\phi_3)}\mathrm{e}^{2\mathrm{i}(\bar{\beta}-\varphi-\alpha_3)}J_2(R\,A_3') + \mathrm{e}^{\mathrm{i}(\phi_3-\phi_2}\mathrm{e}^{-2\mathrm{i}(\bar{\beta}+\alpha_1)}J_2(R\,A_1')\\
+\mathrm{e}^{\mathrm{i}(\phi_3-\phi_1-2\phi_2)}\mathrm{e}^{2\mathrm{i}(\bar{\beta}+\varphi-\alpha_2)}J_2(R\,A_2')\Big]\,.
\end{aligned} \tag{5.6}$$

The R-integrations can now be performed with a special-purpose integrator suited to deal with integrals over oscillating functions. The most canonic choice would be Fast Hankel Transforms, which constitute an analogue to the FFT algorithm where one filters the target function over a Bessel function $J_n(\ell\theta)$ instead of the exponential $\mathrm{e}^{-i\ell\theta}$. This has been beautifully implemented in the FFTLog algorithm (Talman, 1978; Hamilton, 2000). Unfortunately, this requires the integral to be of the form

$$f(\theta) = \int \mathrm{d}\ell\, g(\ell) J_n(\ell\theta)\,; \tag{5.7}$$

since the function we need to integrate depends on both φ and ψ, and these dependencies do not decouple, we can not use this method. We thus opt for the special-purpose integration method derived by Ogata (2005), which is described in Sect. 2.3.2.3. This allows us to solve the R-integration with high numerical accuracy. The subsequent φ and ψ integrations contain only the slowly oscillating complex exponentials and can thus

be solved with any common numerical integration method. For our purposes, we chose an h-adaptive Gauss-Kronod quadrature method (see Sect. 2.3.2.1) implemented in the software CUBATURE[2].

To model the other natural components of the 3pcf, $\Gamma^{(2)}$ and $\Gamma^{(3)}$, we utilise the transformation laws in Eq. (2.126). A careful reader might notice that the natural components we model from the bispectrum assume the orthocenter as the triangle centre, whereas the transformation to aperture mass statistics (Eq. 2.134) defines the centroid as the centre of a triangle. To ensure compatibility with these transformations, we use Eq. (2.125) to transform the modelled 3pcf to the centroid-based ones.

We implement the integration routine on graphics processing units (GPUs) to ensure fast modelling times. With the high level of parallelisation, modelling one bin takes about 5-10 seconds (depending on the triangle configuration). This means that modelling a shear 3pcf in 10^3 bins takes about 2-3 hours. Thus, performing a MCMC-based cosmological parameter analysis, even if nested sampling methods are utilised, is not feasible. Furthermore, estimating a covariance matrix for a data vector of about 8×10^3 is impossible with N-body simulations.

5.1.2.2. Measuring

We use the public code TREECORR (Jarvis et al., 2004) to measure the shear 3pcf. TREECORR is a highly optimised tree-code (as described in Sect. 2.1.2.6) that is written in the fast programming language C++ and allows for a high level of parallelisation, meaning that the 3pcf of a galaxy catalogue can be calculated on several hundreds of central processing units (CPUs) simultaneously. A calculation of the 3pcf on GPUs is not yet possible. Although brute-force calculations of third-order statistics on GPUs outperform CPU-based tree-codes for different applications (Linke et al., 2020), this approach is not feasible in cosmic shear. The abovementioned case calculates the correlation function for galaxy-galaxy-galaxy lensing, which usually has significantly lower galaxy number densities. At higher number densities, the $\mathcal{O}(N \log N)$-scaling of CPU-based tree-codes out-performs the $\mathcal{O}(N^3)$-scaling of a GPU-based brute-force algorithm, even if the latter one is faster for lower source counts.

Unfortunately, the runtime of the tree code heavily depends on the number of bins chosen and scales roughly with b^{-4}, where b is a measure for the bin width. Measuring the shear 3pcf for about 10^7 source galaxies distributed over a $10 \times 10\,\mathrm{deg}^2$ field (like in the *Euclid*-like SLICS described in 3.3.1) in 10^3 bins takes about 1 500 CPUh. Measuring the shear 3pcf for a Stage-III survey is thus very time-consuming but definitely feasible (see App. B). However, getting an accurate covariance model by measuring the 3pcf in multiple simulations can not be done at the current stage.

For the 3pcf we adapt the binning scheme of TREECORR: If we assume $x_1 > x_2 > x_3$ for the side-lengths of a triangle, they define the same triangle via the values $r \in [0, \infty]$, $u \in [0, 1]$ and $v \in [-1, 1]$ by

$$r = x_2, \quad u = \frac{x_3}{x_2}, \quad v = \pm\frac{x_1 - x_2}{x_3} \,. \tag{5.8}$$

[2] https://github.com/stevengj/cubature

Here, the sign of v is positive if the sides of the triangle are oriented counter-clockwise and negative otherwise. We note that this implies $\Gamma^{(i)}(r,u,v) = \Gamma^{(i)*}(r,u,-v)$. This binning scheme has the advantage that it avoids empty bins that would arise when binning in x_1, x_2 and x_3, for example when $x_1 + x_2 < x_3$. Especially if the size of the triangle is binned logarithmically, this would be a major concern. Considering Eq. (2.126) it is straightforward to see that knowledge of all natural components $\Gamma^{(i)}(x_1,x_2,x_3)$ for all $x_1 > x_2 > x_3$ is already enough to recover the full information content of the shear 3pcf; for all other orders of x_1, x_2, and x_3, we can simply convert the natural components into each other.

From here, unless otherwise stated, we always bin the shear 3pcf in r, u, and v. We always choose the same number of bins for each of the variables; u and v are binned linearly between 0 and 1 (with complex conjugations of the $\Gamma^{(i)}$ accounting for the case of negative v), and r is binned logarithmically between a minimum and a maximum scale. If we state that we bin the 3pcf in 10^3 bins from $0.'1$ to $120'$, for example, we mean that we take 10 bins for each r, u, and v, and that the bins in r are logarithmic between $0.'1$ and $120'$.

5.1.2.3. Validation

Validating the integration routine As the integration detailed in Eqs. (5.5) and (5.6) is numerically quite difficult, we want to independently test this part of the modelling pipeline. Inspired by the calculations in Shi et al. (2011), we define the three-point correlation function of the deflection potential as

$$\langle \Psi(\boldsymbol{X})\Psi(\boldsymbol{Y})\Psi(\boldsymbol{Z})\rangle = \frac{1}{8\alpha^3}\mathrm{e}^{-\alpha\left[(\boldsymbol{X}-\boldsymbol{Y})^2+(\boldsymbol{Y}-\boldsymbol{Z})^2+(\boldsymbol{X}-\boldsymbol{Y})^2\right]} . \tag{5.9}$$

We then define

$$\partial_X = \partial_{X_1} + \mathrm{i}\partial_{X_2} , \quad \nabla_X^2 = \partial_X \partial_X^* , \tag{5.10}$$

and, using the relations

$$\langle \kappa(\boldsymbol{X})\kappa(\boldsymbol{Y})\kappa(\boldsymbol{Z})\rangle = \left(\frac{1}{2}\nabla_X^2\right)\left(\frac{1}{2}\nabla_Y^2\right)\left(\frac{1}{2}\nabla_Z^2\right)\langle \Psi(\boldsymbol{X})\Psi(\boldsymbol{Y})\Psi(\boldsymbol{Z})\rangle , \tag{5.11}$$

$$\langle \gamma_\mathrm{c}(\boldsymbol{X})\gamma_\mathrm{c}(\boldsymbol{Y})\gamma_\mathrm{c}(\boldsymbol{Z})\rangle = \left(\frac{1}{2}\partial_X^2\right)\left(\frac{1}{2}\partial_Y^2\right)\left(\frac{1}{2}\partial_Z^2\right)\langle \Psi(\boldsymbol{X})\Psi(\boldsymbol{Y})\Psi(\boldsymbol{Z})\rangle , \tag{5.12}$$

$$\langle \gamma_\mathrm{c}(\boldsymbol{X})\gamma_\mathrm{c}(\boldsymbol{Y})\gamma_\mathrm{c}^*(\boldsymbol{Z})\rangle = \left(\frac{1}{2}\partial_X^2\right)\left(\frac{1}{2}\partial_Y^2\right)\left(\frac{1}{2}\partial_Z^{*2}\right)\langle \Psi(\boldsymbol{X})\Psi(\boldsymbol{Y})\Psi(\boldsymbol{Z})\rangle , \tag{5.13}$$

we can analytically calculate both the shear 3pcf and the convergence bispectrum. When we define $\boldsymbol{x} = \boldsymbol{X} - \boldsymbol{Z}$ and $\boldsymbol{y} = \boldsymbol{Y} - \boldsymbol{Z}$, we can calculate:

$$\begin{aligned}\langle \tilde{\kappa}\tilde{\kappa}\tilde{\kappa}\rangle(\boldsymbol{\ell_1},\boldsymbol{\ell_2},\boldsymbol{\ell_3}) &= -\frac{\pi^4}{6\alpha^5}\ell_1^2\ell_2^2\ell_3^2\,\delta_\mathrm{D}(\boldsymbol{\ell_1}+\boldsymbol{\ell_2}+\boldsymbol{\ell_3})\mathrm{e}^{-(\ell_1^2+\ell_2^2+\ell_3^2)/12\alpha} \\ &= -\frac{\pi^4}{6\alpha^5}\ell_1^2\ell_2^2(\ell_1^2+\ell_2^2+2\boldsymbol{\ell_1}\cdot\boldsymbol{\ell_2})\,\delta_\mathrm{D}(\boldsymbol{\ell_1}+\boldsymbol{\ell_2}+\boldsymbol{\ell_3})\mathrm{e}^{-(\ell_1^2+\ell_2^2+\boldsymbol{\ell_1}\cdot\boldsymbol{\ell_2})/6\alpha} \end{aligned} \tag{5.14}$$

$$= (2\pi)^2 B_\kappa(\ell_1,\ell_2,\ell_3)\delta_\mathrm{D}(\boldsymbol{\ell_1}+\boldsymbol{\ell_2}+\boldsymbol{\ell_3}) , \tag{5.15}$$

$$\langle \gamma_\mathrm{c}\gamma_\mathrm{c}\gamma_\mathrm{c}\rangle(\boldsymbol{x},\boldsymbol{y}) \overset{(*)}{=} \alpha^3\left[(\boldsymbol{x}-2\boldsymbol{y})(\boldsymbol{x}+\boldsymbol{y})(\boldsymbol{y}-2\boldsymbol{x})\right]^2\,\mathrm{e}^{-2\alpha(x^2+y^2-\boldsymbol{x}\cdot\boldsymbol{y})} . \tag{5.16}$$

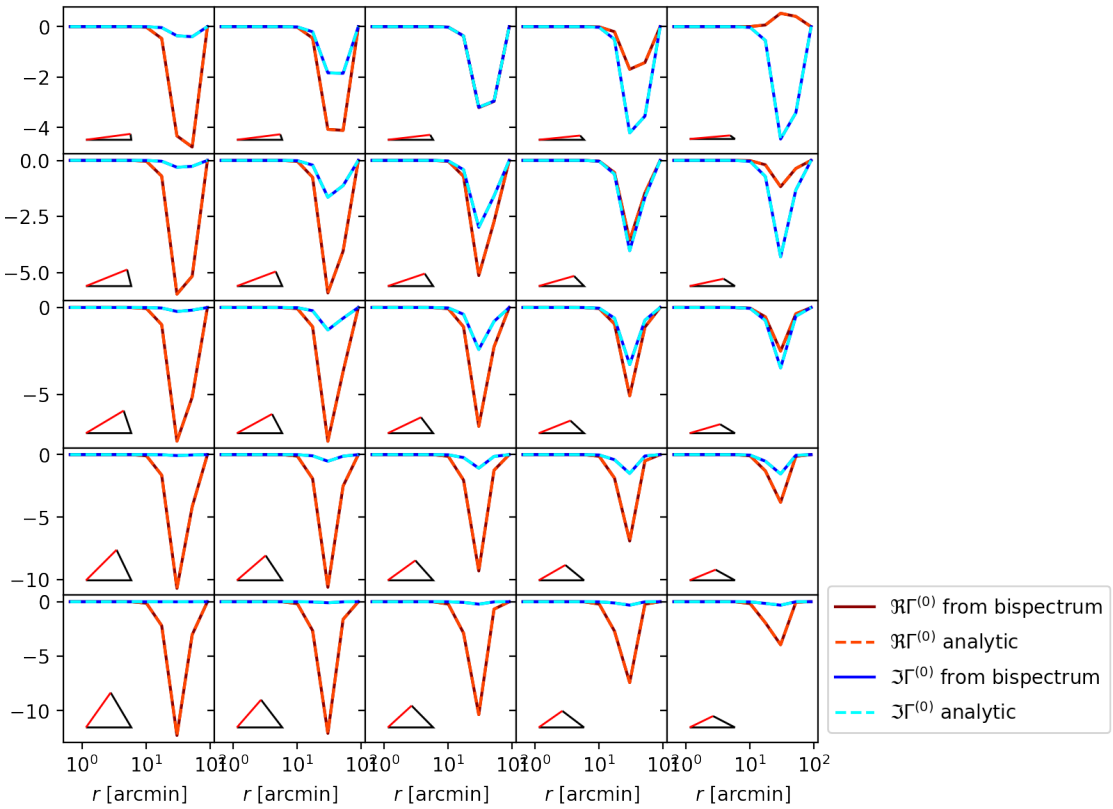


Figure 5.3. – The real (red) and imaginary (blue) parts of the first natural component of the shear three-point correlation functions $\Gamma^{(0)}$. Each plot panel corresponds to one fixed triangle shape shown in the bottom-right corner; the x-axis represents the length of the red triangle shape. We compare the integration of the analytic bispectrum model (Eq. 5.17, dark solid) and the analytic solution (Eq. 5.16, light dashed). From Heydenreich et al. (2022b).

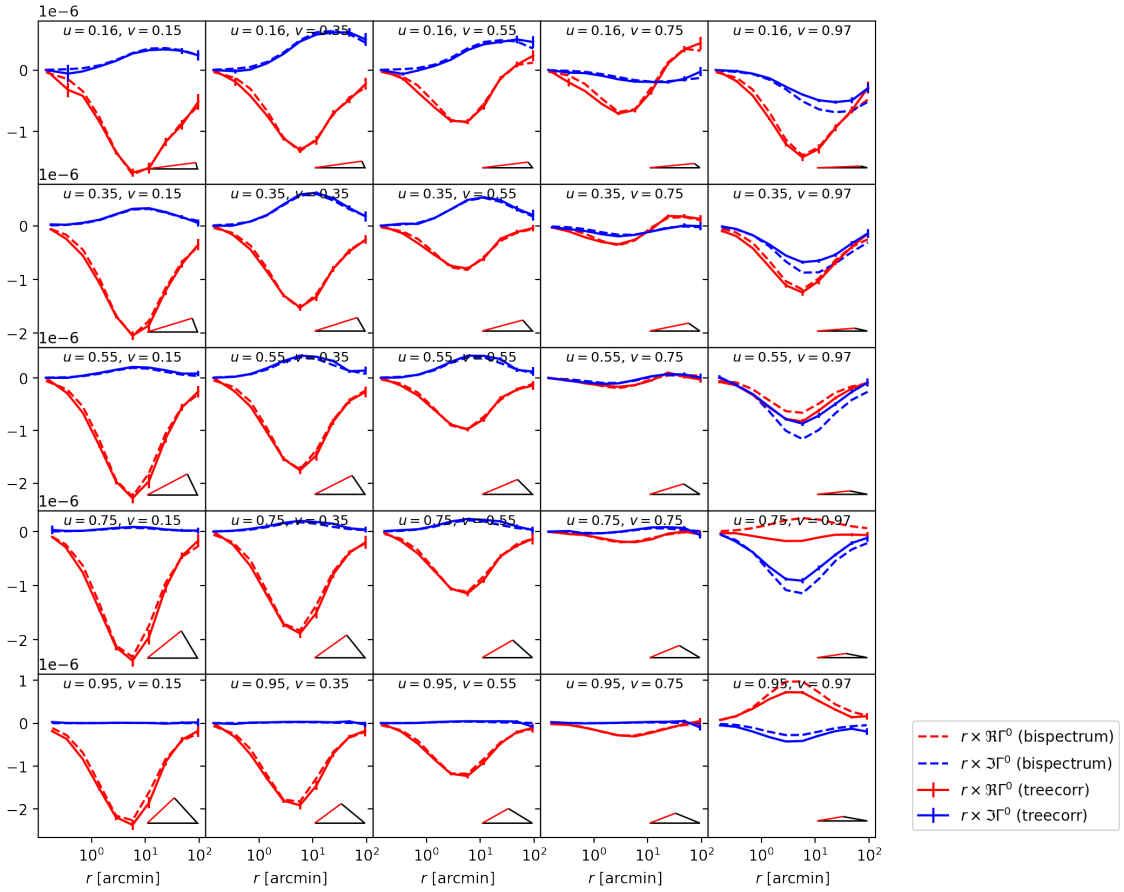


Figure 5.4. – The first natural component $\Gamma^{(0)}$ of the shear three-point correlation functions modelled from BiHalofit compared to measurements from the MS that were extracted using treecorr. Each panel corresponds to one fixed triangle combination; the u and v values are listed at the top, and the corresponding shape of the triangle is shown on the bottom right. The x-axis corresponds to the length of the red side of the triangle. We show both the real part (red) and the imaginary part (blue) for the model (dashed line) and the simulations (solid line). The error bars denote the error on the mean of the 64 lines-of-sight of the MS. From Heydenreich et al. (2022b).

In the last equation, marked by $(*)$, the variables $\boldsymbol{x}, \boldsymbol{y}$ are interpreted as complex numbers $\boldsymbol{x} = x_1 + \mathrm{i}x_2$; for their scalar product, $\boldsymbol{x}\cdot\boldsymbol{y} = x_1y_1 + x_2y_2$ holds. The equation for $\langle\gamma_\mathrm{c}\gamma_\mathrm{c}\gamma_\mathrm{c}^*\rangle$ can be computed via Mathematica and is too long to denote here. We can now set

$$b(\ell_1,\ell_2,\varphi) = -\frac{\pi^2\ell_1^2\ell_2^2(\ell_1^2+\ell_2^2+2\ell_1\ell_2\cos\varphi)}{24\alpha^5}\mathrm{e}^{-(\ell_1^2+\ell_2^2+\ell_1\ell_2\cos\varphi)/6\alpha}\,. \tag{5.17}$$

We then transform the Cartesian shears in Eq. (5.16) to the orthocenters using Eq. (2.128). Using the definitions $\Gamma^{(0)} = \langle\gamma\gamma\gamma\rangle$ and $\Gamma^{(3)} = \langle\gamma\gamma\gamma^*\rangle$ (see Eq. 2.123), we can validate our integration routine. We show the results in Fig. 5.3 and verify that the integration routine is accurate to the sub-per cent level, which is more than enough given that we have a 10% uncertainty on our bispectrum model.

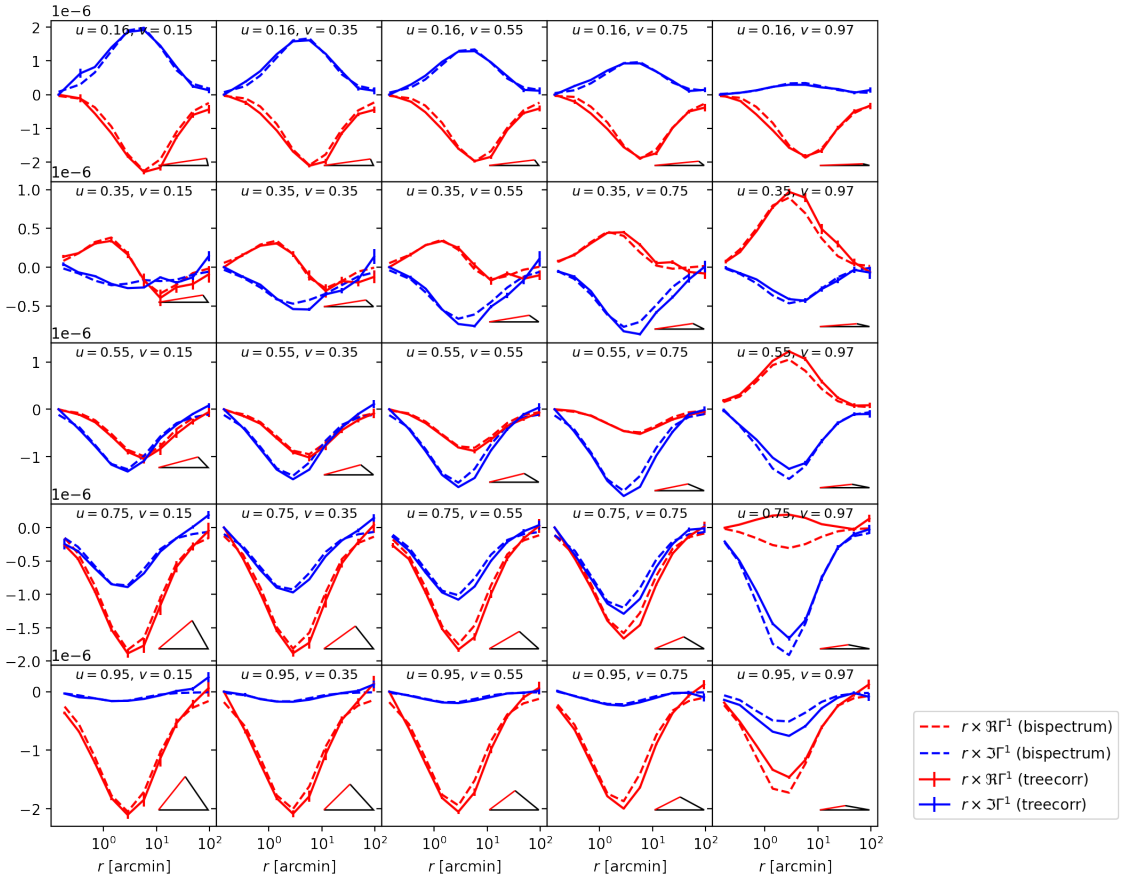


Figure 5.5. – Same as Fig. 5.4, but for the second natural component $\Gamma^{(1)}$. From Heydenreich et al. (2022b).

Validating the 3pcf model We again use the MS to validate our model by measuring the shear 3pcf in 10^3 bins between $0.'1$ and $120'$ at redshift $z = 1$. We then model the 3pcf at the bin centers and show the results in Fig. 5.4. We can see a good agreement over a large range of scales. We were especially surprised by the level of agreement down to sub-arcminute scales, which roughly correspond to wavenumbers of $\ell \gtrsim 10^4$. In practice, these scales can probably not be used in a real data analysis due to contamination by baryonic effects, but it is still encouraging to see this level of agreement. Only for highly degenerate triangles (see the right column of Fig. 5.4) do we notice that the results do not agree; whether this is due to a breakdown of the bispectrum model or the tree-code used to measure the 3pcf remains to be seen. We see a similar level of agreement for the other three natural components $\Gamma^{(1)}, \Gamma^{(2)}$ and $\Gamma^{(3)}$ (for an example, see Fig. 5.5).

5.1.3. Third-order aperture statistics

5.1.3.1. Modelling

We model the third-order aperture statistics using Eq. (2.133). In this case, the filter functions $\tilde{u}$ are smooth and do not oscillate, which means that the integral is relatively simple to solve numerically, and we do not need to implement any special-purpose integration routines. We thus again utilise a GPU-based implementation of the h-adaptive cubature, which allows us to model the third-order aperture statistics within a few seconds. As the number of bins required for these statistics is much lower (about 20-100 in contrast to the 1000-10 000 for the 3pcf), the modelling times become feasible for a cosmological parameter analysis.

We also model the second-order aperture statistics from Eq. (2.115) to compare their constraining powers, where we perform a Limber integration (Eq. 2.105) of the revised HALOFIT model for the matter power spectrum (Takahashi et al., 2012).

5.1.3.2. Measuring

We have already discussed that one of the advantages of aperture mass statistics is that there is a way to measure them quickly in simulations and another one to measure them in surveys with complex geometries. Here, we discuss our implementations of the different measuring strategies.

Direct measurement While direct measurement of the aperture statistics is not feasible in real data, it can be efficiently performed in simulations. To calculate the aperture mass map, we utilise the estimator:

$$\widehat{\mathcal{M}}_{\mathrm{ap}}(\boldsymbol{\vartheta};\theta) + \mathrm{i}\widehat{\mathcal{M}}_{\perp}(\boldsymbol{\vartheta};\theta) = \frac{1}{n_{\mathrm{gal}}} \sum_i Q_\theta(|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_i|) \left(\varepsilon_{\mathrm{t},i} + \mathrm{i}\varepsilon_{\times,i}\right) \ . \tag{5.18}$$

Here, $\varepsilon_{\mathrm{t}/\times}$ are the tangential and cross component of the galaxy ellipticities, defined in analogue to Eq. (2.90), and $\boldsymbol{\vartheta}_i$ are their respective positions; n_{gal} can be defined as the global number density of galaxies (Bartelmann and Schneider, 2001) or as the number density of galaxies within the aperture radius (Martinet et al., 2018). For this work, we define n_{gal} as the number of galaxies weighted by the Q-filter function:

$$n_{\mathrm{gal}} = \sum_i Q_\theta(|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_i|) \ . \tag{5.19}$$

We test all three definitions of n_{gal} and find that setting n_{gal} as the number density within the aperture radius or the one weighted by the Q-filter function induces sub-per cent differences on the third-order aperture masses $\left\langle \mathcal{M}_{\mathrm{ap}}^3 \right\rangle$ of a single map. However, setting n_{gal} as the global galaxy density can induce differences of about 5% in $\left\langle \mathcal{M}_{\mathrm{ap}}^3 \right\rangle$. In both cases, these differences diminish when we average over several realisations of aperture mass maps. A similar estimator of the aperture mass map can be constructed for a convergence field by transforming the integral in Eq. (2.91) into a sum.

When observing the estimator in Eq. (5.18), we see that it can be written as a convolution. We thus place the observed galaxy ellipticities on a grid and, similar to Sect. 4.1,

calculate the aperture mass map by utilising the convolution theorem. The same approach can be applied to calculate the aperture mass maps from a convergence field. To avoid boundary effects, we always cut off a slice of width $4\theta_{\rm ap}$ around the border of an aperture mass map. In the full-sky T17, we utilise the HEALPY function SMOOTHING to calculate the convolution on a sphere. The smoothing kernel can be calculated from the corresponding U- and Q-filters using the function BEAM2BL. In this case, boundary effects are not present.

This method allows us to quickly and efficiently calculate the aperture mass maps, and thus the third-order aperture statistics, in simulations. In particular, we can estimate the covariance of third-order aperture statistics from numerical simulations[3].

Conversion from shear 3pcf To measure third-order aperture statistics in a real survey with boundaries and masked regions, we want to first measure the shear 3pcf and then convert them into aperture statistics. This avoids any potential issues that the survey geometry might cause. This procedure has already been implemented into TREECORR. The conversion is achieved by performing the integral in Eq. (2.134). As we have already discussed at the end of Sect. 5.1.2.2, we can fully reconstruct both $\Gamma^{(0)}$ and $\Gamma^{(3)}$ from the measurements performed by TREECORR. Performing the appropriate permutations[4], we can thus transform $\Gamma^{(1)}$ and $\Gamma^{(2)}$ into $\Gamma^{(3)}$ and calculate the determinant of the transformation matrix $\left|\det \frac{\mathrm{d}\{y_1,y_2,\psi\}}{\mathrm{d}\{r,u,v\}}\right|$. The integration is then performed by taking the sum of all values of the 3pcf multiplied by the determinant of the transformation matrix and the value of the filter functions $F_{1,2}$ at the bin centres. This is certainly not an optimal procedure, and we discuss potential improvements in Sect. 5.3.

5.1.3.3. Validation

Binning choices of the 3pcf First, we validate the conversion from shear 3pcf to third-order aperture masses. We want to determine the minimum and maximum scale and the number of bins in which we need to measure the 3pcf to get an unbiased estimate of the third-order aperture statistics. Furthermore, we want to verify that the E- and B-mode leakage for third-order aperture statistics is not too severe. In order to do this, we model the shear 3pcf between $\theta_{\rm min} = 0.'1$ and $\theta_{\rm max} = 120'$ with various binning choices and use the method outlined above to convert them to third-order aperture statistics. We then compare these to third-order aperture statistics that we model directly from the convergence bispectrum. Since both methods model the third-order aperture statistics from the same bispectrum, their results should agree. We show the results of this test in Fig. 5.6. We notice that the conversion tends to fail particularly for configurations of aperture radii where one radius is much larger than the other two. For roughly equal-scale aperture radii, a measurement of the 3pcf in 10^3 bins appears sufficient, but when the aperture radii differ by a factor of four or more, one has to resort to measuring the 3pcf in 15^3 bins. The increase to 20^3 bins appears to yield only a marginal improvement; we

[3] There is an issue when one calculates the covariance of aperture statistics from simulations that follow a different geometry than the actual survey. We elaborate on this in Sect. 5.2

[4] The previously mentioned bug in TREECORR did not correctly perform these permutations before calculating the transformation matrix.

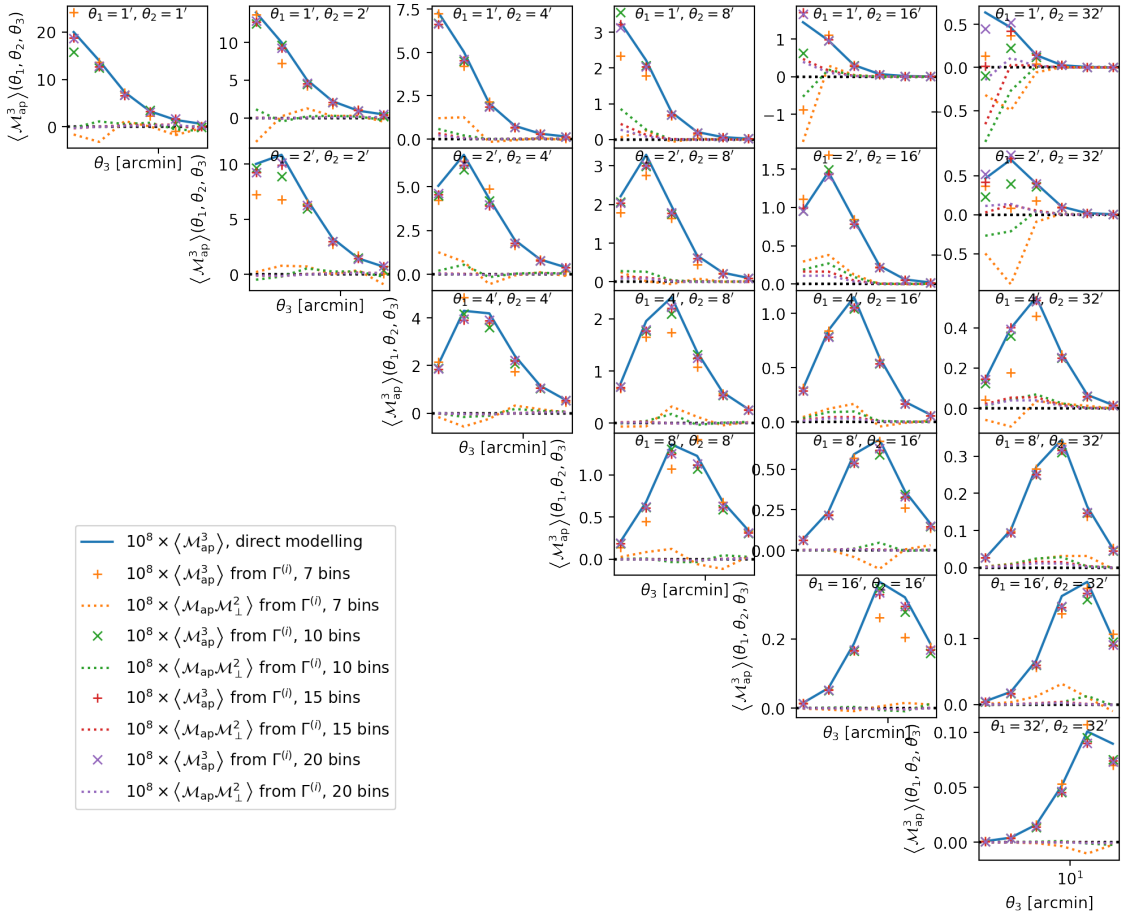


Figure 5.6. – We compare the directly modelled third-order aperture statistics $\langle \mathcal{M}_{\rm ap}^3 \rangle (\theta_1, \theta_2, \theta_3)$ (blue line) to the ones that we convert from the modelled shear 3pcf (coloured crosses). In each panel, the x-axis describes the value of θ_3; θ_1 is constant in each row and θ_2 is constant in each column. The yellow, green, red and purple crosses arise from the shear 3pcf that have been binned in $7^3, 10^3, 15^3$ and 20^3 bins in r, u and v, respectively. The dotted lines denote the corresponding B-modes.

decide that measurement in 15^3 bis is a good compromise between speed and accuracy. We further note that (unsurprisingly) the B-mode leakage is highest at the points where the conversion tends to fail but well below 10% in almost all cases. As our bispectrum model is only accurate to 10-20%, we deem this level of leakage acceptable at the current stage. Our results extend upon the findings of Shi et al. (2014), who performed a similar test for a simplified bispectrum model and equal-scale aperture radii and found per cent-level leakage for all scales larger than $10\,\theta_{\rm min}$. From a future analysis, we plan to exclude all combinations of aperture radii where the B-mode leakage exceeds 10%.

We further investigate how the shear 3pcf depends on r, u and v and show an interesting case in Fig. 5.7. We can see there that the dependence of the 3pcf on the parameter r is relatively well-behaved, whereas it strongly varies as a function of u and v, especially for $u \approx 0$ and $|v| \approx 1$, meaning that a fine binning in the latter two parameters is essential.

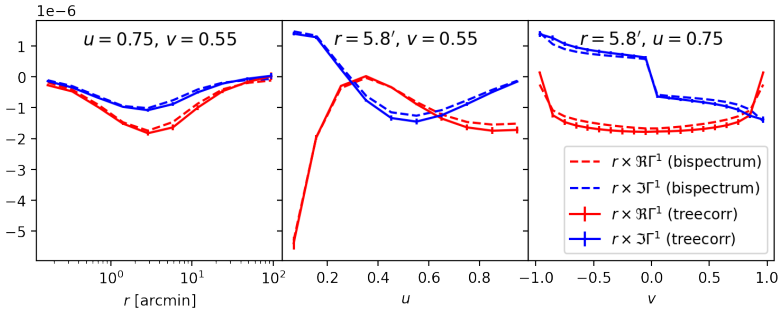


Figure 5.7. – The second natural component $\Gamma^{(1)}$ of the 3pcf as a function of its three parameters r (left), u (middle) and v (right). We note that since $\Gamma^{(i)}(r, u, v) = \Gamma^{*(i)}(r, u, -v)$ holds, the jump in $\mathrm{Im}\,\Gamma^{(1)}$ at $v = 0$ is expected.

This contrasts the binning choices made in Secco et al. (2022a), who opted to measure the 3pcf in 55 fine r-bins and only 10 bins in each u and v.

Comparison to N-body simulations We validate our modelling pipeline with the MS in Fig. 5.8, where we compare aperture masses obtained from both direct measurement and conversion from shear 3pcf to our model predictions and find a good overall agreement between all three. As expected, the conversion from the 3pcf to aperture mass statistics fails in the top-right corner, where two aperture radii are small and the third one is large. In that region, we also observe a significant amount of B-modes. We also note that the error bars on the direct measurement are larger for most points since we have to cut off a significant amount of the aperture mass maps to avoid boundary effects.

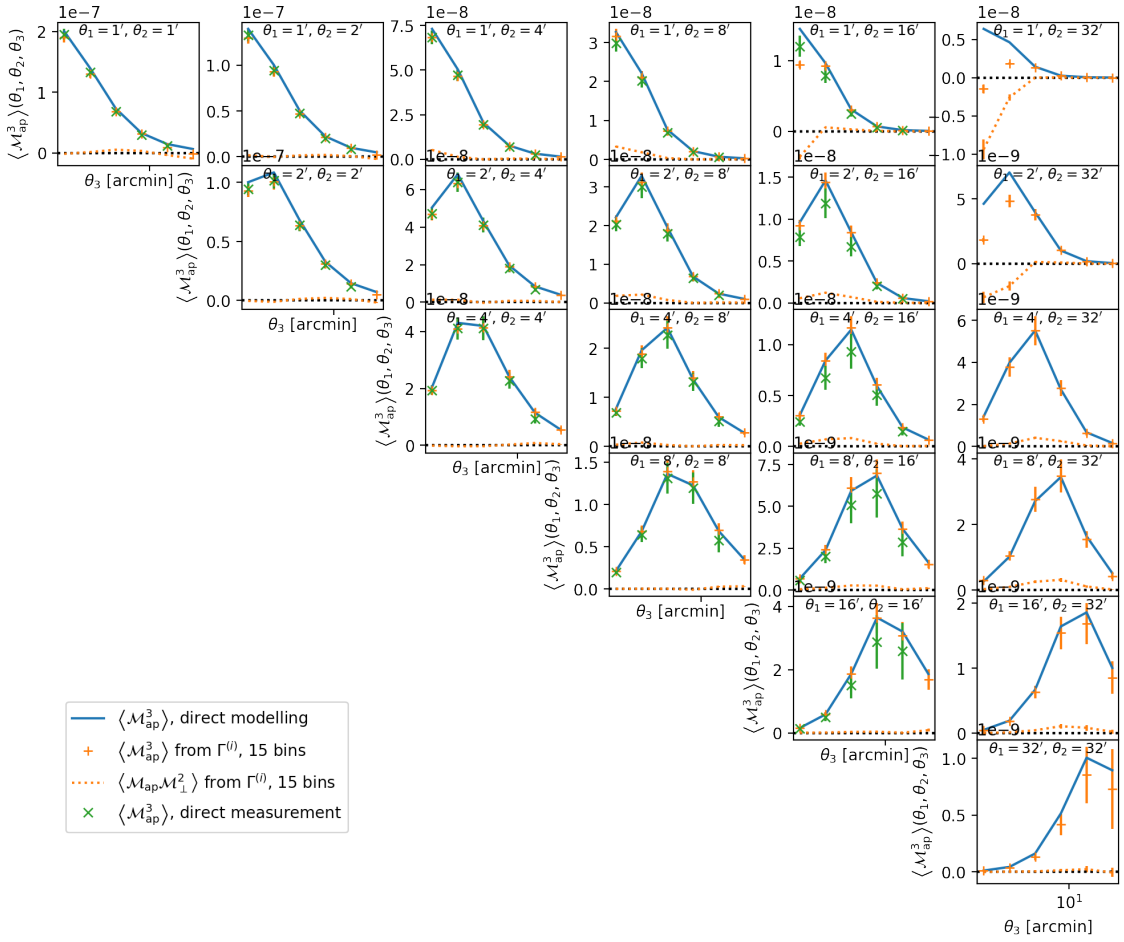


Figure 5.8. – The third-order aperture masses from our modelling pipeline (blue) compared to the ones measured in the MS. The direct measurements are shown in green, and the measurements converted from shear 3pcf are shown in orange. Both statistics were computed from 32 lines of sight; the error bars represent the error on the mean. We note that for the largest angle of $32'$, a direct measurement could not be obtained due to the limited size of the light cone. From Heydenreich et al. (2022b).

5.1.4. Cosmological parameter estimation

Having validated all aspects of our modelling pipeline, we aim to assess the constraining power of third-order aperture statistics. For this study, we focus on three questions. Firstly, how much information do we lose when we convert shear 3pcf to third-order aperture statistics? Secondly, how much information do we gain when we combine second- and third-order aperture statistics compared to a pure second-order analysis? And thirdly, what is the information content of equal-scale aperture mass statistics compared to utilising all combinations of aperture radii?

We perform a series of cosmological parameter inferences on mock data to answer these questions using a MCMC framework. To speed up the modelling pipeline, we create 7500 nodes in a four-dimensional Latin hypercube describing the cosmological parameters $\Omega_\mathrm{m}, S_8, w_0$ and h in a flat wCDM cosmology; all other cosmological parameters are fixed to the fiducial values of the SLICS. We then calculate the third-order aperture statistics for all combinations of aperture radii of $4', 8', 16'$ and $32'$ at these nodes. Afterwards, we train the CosmoPower emulator (Spurio Mancini et al., 2022) on 6500 of the training nodes and use the remaining 1000 to validate its performance. As can be seen in Fig. 5.9, the uncertainty of the emulator is far below the uncertainty of our bispectrum model and thus completely negligible for us.

As our covariance model is not completely developed yet, we extract the covariances for our analysis from N-body simulations. To account for that fact, we adapt the likelihood function of Percival et al. (2022), which we introduce in Sect. 2.3.3.1.

5.1.4.1. Shear three-point correlation functions compared to third-order aperture statistics

As established before, we can not perform a cosmological parameter inference with shear 3pcf. We thus model the shear 3pcf in 10^3 bins between $0.'1$ and $100'$ at 500 of the 7500 training nodes and perform a PCA to compress the 8 000 entries of the data vector to its 40 most relevant principal components (similar to the data compression method we tested in Sect. 4.3.3). We then perform the same PCA on data vectors that we measure from 200 tiles of the *Euclid*-like SLICS to estimate a covariance matrix. We train the CosmoPower emulator on the 40 principal components at all 500 training nodes[5].

We compare the parameter constraints of the principal components of the 3pcf to the ones from third-order aperture statistics, where we use all 927 available tiles of the SLICS to estimate the covariance. In both cases, we model a data vector at the fiducial SLICS cosmology with the same pipeline for the MCMC, as we do not expect to recover the input cosmology in a *Euclid*-like setting. As we can see in Fig. 5.10, there is only a marginal difference in constraining power. Of course, this is not a rigorous comparison: The shear 3pcf and third-order aperture statistics use different scales, and there is certainly some information loss that arises from the PCA. Furthermore, as we cut off the boundaries of the aperture masses, we discard a part of the data. All things considered, we now have

[5] An independent validation test, where we only train the emulator on 400 nodes and validate on the remaining 100 ones, shows that the performance of the emulator is worse than for third-order aperture statistics but still satisfactory.

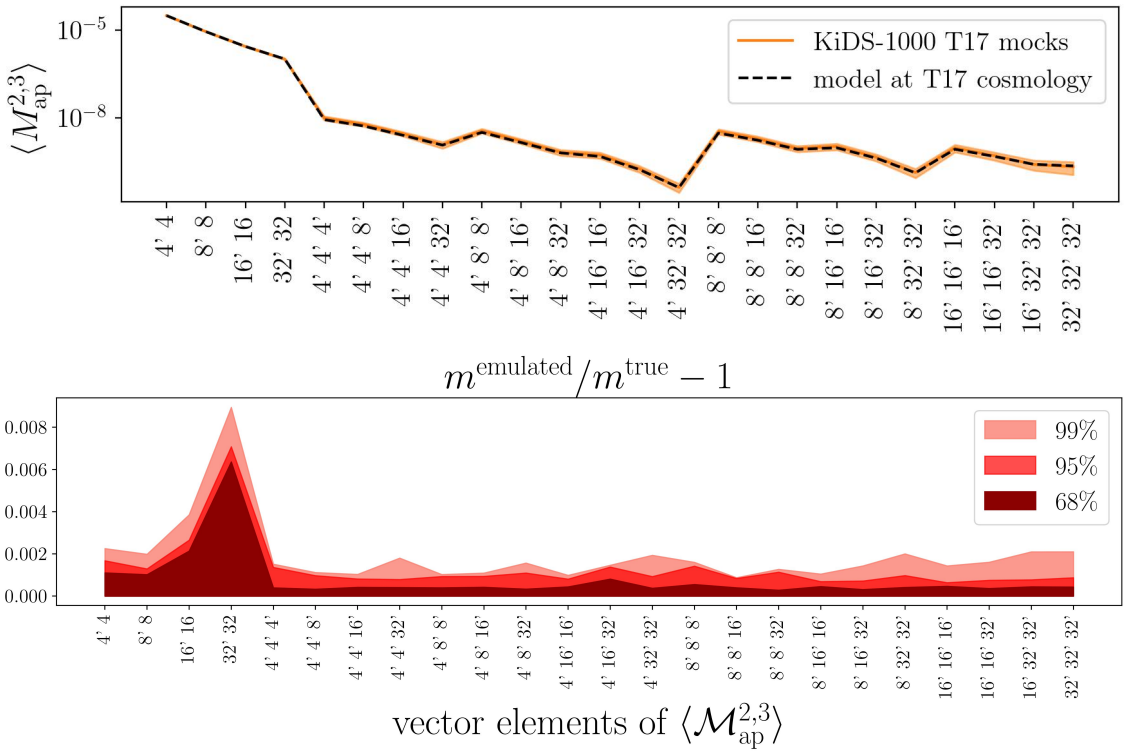


Figure 5.9. – The top panel shows the data vector for a combination of second- and third-order aperture statistics $\langle \mathcal{M}_{\rm ap}^{2,3} \rangle$. The dashed black line shows our model at the T17 cosmology, the orange band shows the measurement in the simulations and its standard deviation for a KiDS-1000-like setup. The bottom panel shows the fractional error of the emulated second- and third-order aperture statistics in the same setup.

some confidence that we do not face a severe information loss when converting shear 3pcf to third-order aperture statistics.

5.1.4.2. Constraining power of a joint analysis

We forecast the constraining power of a joint analysis between second- and third-order shear statistics in a Stage-III survey[6]. We extract a covariance matrix for the second- and third-order aperture statistics from the T17 simulations with a resolution of NSIDE = 4096, corresponding to a pixel size of $0\farcm74^2$, where we employ a KiDS-1000-like source redshift distribution and cut out 18 independent tiles of $\approx 860\,\mathrm{deg}^2$ from each full-sky realisation. This yields a total of 1944 independent realisations from which we extract our covariance matrix. We then estimate our target data vector from one full-sky realisation of the T17 with NSIDE = 8192, as the lower-resolution maps are inconsistent with our model at small scales. The data vector, its uncertainty, and our model can be seen in Fig. 5.9.

In Fig. 5.11 we find that a joint analysis significantly improves the constraining power on all cosmological parameters with respect to an analysis with only second-order statistics.

[6]For a Stage-IV survey, our bispectrum model is by far not accurate enough, so we refrain from making any quantitative predictions for that.

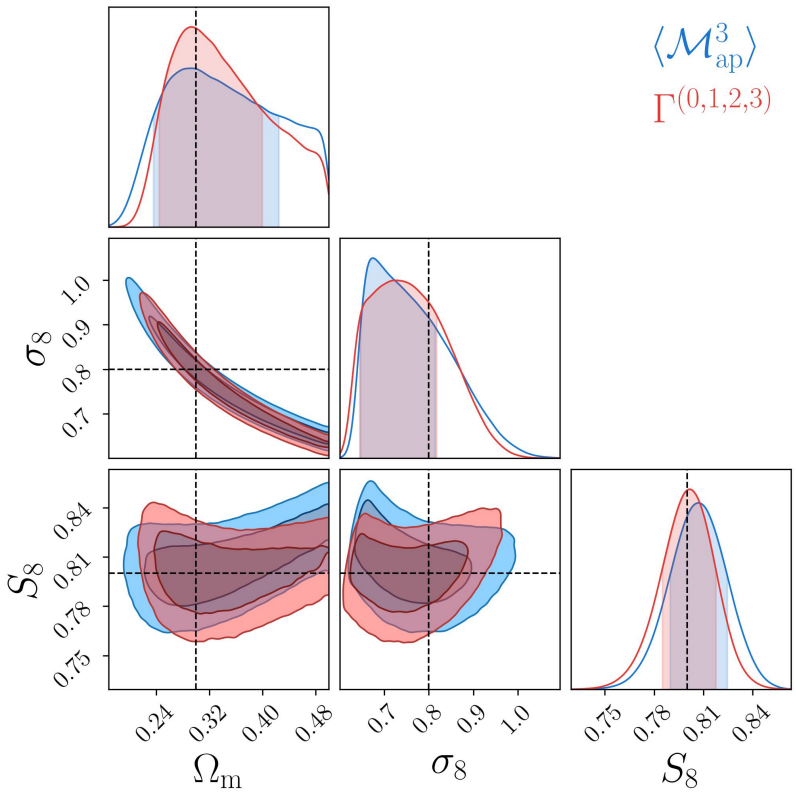


Figure 5.10. – A comparison of the parameter constraints achieved by the principal components of the shear 3pcf (red) and third-order aperture statistics (blue) for a covariance from the *Euclid*-like SLICS. The dashed line indicates the cosmology of the data vector that was used in the MCMC.

In particular, the uncertainty on S_8 is reduced by 43%, and the figure-of-merit (Albrecht et al., 2006) in the $\Omega_{\rm m}$-σ_8 plane is increased by a factor of almost 6. We can further see that the second- and third-order shear statistics exhibit different degeneracy directions between $\Omega_{\rm m}$ and σ_8 (or $\Omega_{\rm m}$ and S_8). This degeneracy breaking also yields substantial improvements on the constraints of $\Omega_{\rm m}$ and σ_8 (more than 68% and 54%, respectively). Lastly, we note that the input cosmology is well within the 1σ region of the parameter constraints for all three statistics, which gives reason to believe that our model for third-order aperture statistics is accurate enough for an analysis of a Stage-III survey.

We then compare the constraining power of a joint analysis with only equal-scale third-order aperture statistics to one where we use all combinations of filter radii in Fig. 5.12 and find that the equal-scale aperture mass statistics $\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle(\theta, \theta, \theta)$ already possess the majority of the information content of the third-order shear signal. Considering the results from Kilbinger and Schneider (2005), this is a bit surprising, although it is in line with the findings of Fu et al. (2014). In particular, this means that one can significantly speed up a cosmological parameter analysis by measuring the shear 3pcf in 10^3 bins (which yields an accurate conversion for equal-scale aperture radii) and only modelling equal-scale aperture radii while only suffering a minor loss of information content. We note that these findings might change when additional parameters (either cosmological or nuisance)

Table 5.1. – Marginalised one-dimensional parameter constraints of Fig. 5.11. For $\langle \mathcal{M}_{\rm ap}^2 \rangle$ and $\langle \mathcal{M}_{\rm ap}^3 \rangle$ we do not cite upper limits on $\Omega_{\rm m}$ or lower limits on σ_8 as they are dominated by the prior.

parameter	$\langle \mathcal{M}_{\rm ap}^2 \rangle$	$\langle \mathcal{M}_{\rm ap}^3 \rangle$	$\langle \mathcal{M}_{\rm ap}^{2,3} \rangle$
$\Omega_{\rm m}$	$0.294_{-0.059}$	$0.229_{-0.048}$	$0.26^{+0.041}_{-0.04}$
σ_8	$0.671^{+0.155}$	$0.603^{+0.287}$	$0.842^{+0.075}_{-0.074}$
S_8	$0.813^{+0.023}_{-0.024}$	$0.786^{+0.022}_{-0.041}$	$0.792^{+0.017}_{-0.019}$

are introduced in the MCMC. In that case, the equal-scale aperture masses will suffer from some degeneracies that the aperture mass statistics containing all combinations of filter radii might be able to break.

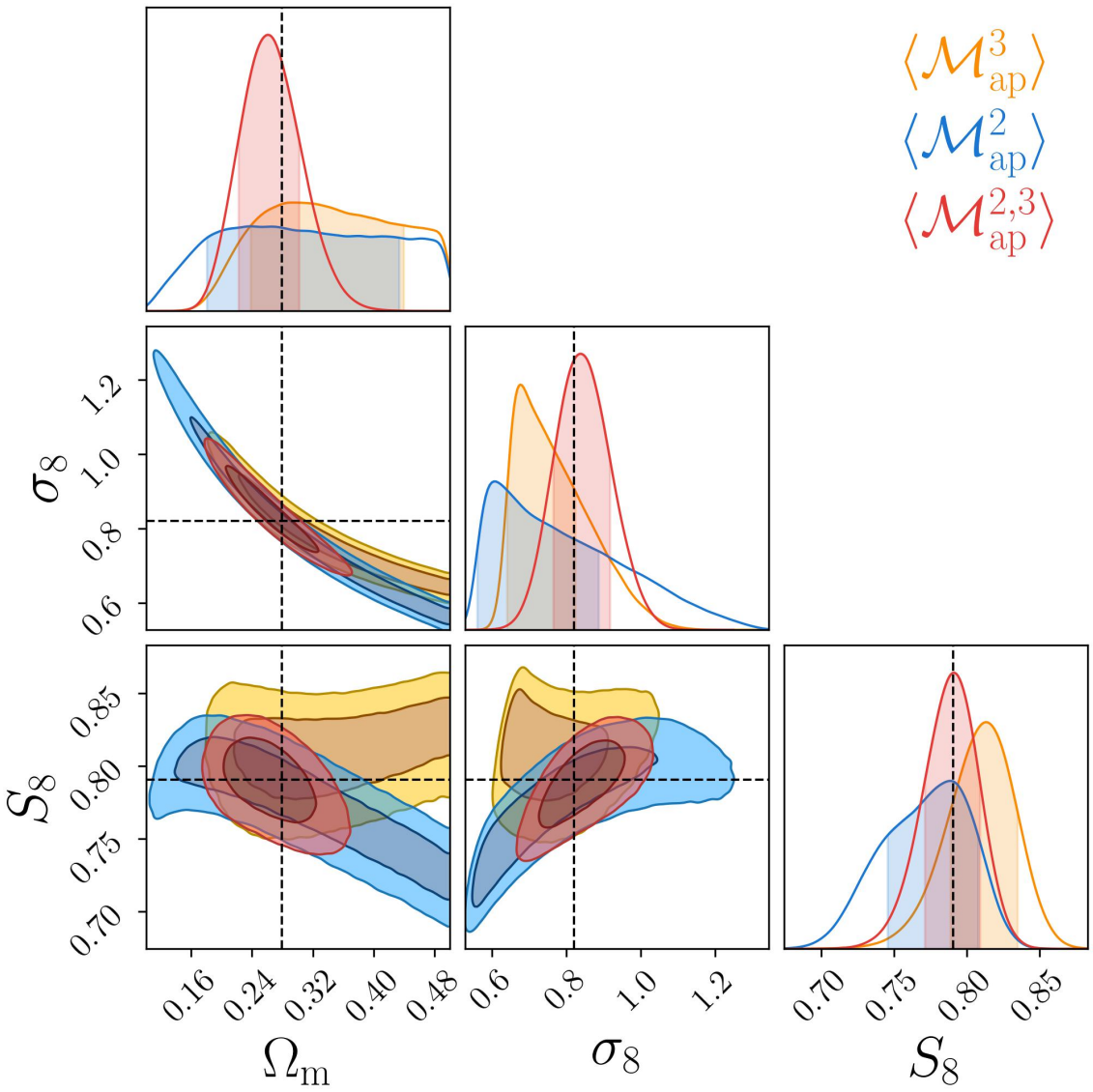


Figure 5.11. – The posterior distributions of second- (blue) and third-order (yellow) aperture statistics with a joint analysis (red) within a KiDS-1000-like setup. We estimate the data vector and covariance matrix with the T17 simulations, and we use filter scales of $(4, 8, 16, 32)$ arcmin. The Hubble parameter h and the dark energy EoS are fixed to the T17 values. The marginalised posteriors can be found in Tab. 5.1. The dashed line indicates the cosmology of the data vector that was used in the MCMC. Adapted from Heydenreich et al. (2022b).

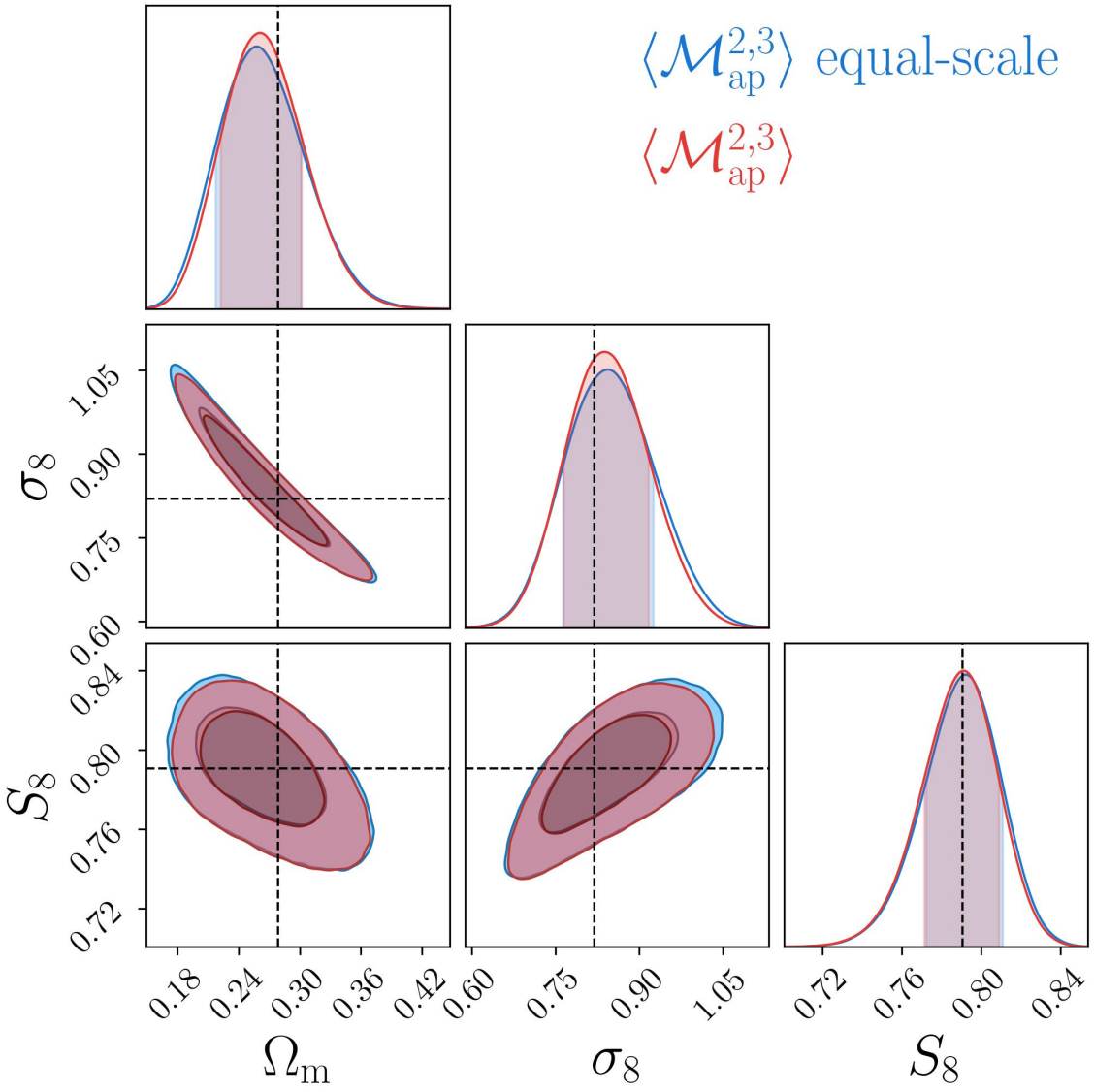


Figure 5.12. – We compare the parameter constraints of a joint analysis with equal-scale third-order aperture statistics (blue) to an analysis where all combinations of filter radii are used (red). We estimate the data vector and covariance matrix with the T17 simulations, and we use filter scales of $(4, 8, 16, 32)$ arcmin. The Hubble parameter h and the dark energy EoS are fixed to the T17 values. The dashed line indicates the cosmology of the data vector that was used in the MCMC. Adapted from Heydenreich et al. (2022b).

5.2. Covariance of third-order aperture statistics

In this ongoing project led by Laila Linke, we model the covariance of the third-order aperture statistics to prepare for a tomographic analysis. The findings of this work will be published soon (Linke et al. in prep.).

In this Section, Laila Linke developed and implemented the covariance model for third-order aperture statistics. Peter Schneider developed the strategy to measure the individual terms of the covariance model from simulations. I was responsible for measuring the full covariance and the single terms of the covariance model in N-body simulations and Gaussian and lognormal random fields. Again, many results from this section originated after discussions among the collaborators, so perfectly disentangling the contributions is impossible.

Throughout this section, we assume that the third-order statistics $\left\langle \mathcal{M}_{\mathrm{ap}}^3 \right\rangle$ are estimated from a convergence field κ, which has N_{pix} pixels of size A_{pix}, and spans an area $A = N_{\mathrm{pix}}\, A_{\mathrm{pix}}$, via the estimator

$$\widehat{\mathcal{M}}_{\mathrm{ap}}^3(\theta_1,\theta_2,\theta_3) \equiv \frac{1}{N_{\mathrm{pix}}} \sum_{i=1}^{N_{\mathrm{pix}}} \sum_{jkl}^{N_{\mathrm{pix}}} A_{\mathrm{pix}}^3\, U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_j|)\, \kappa(\boldsymbol{\vartheta}_j)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, \kappa(\boldsymbol{\vartheta}_k) \times U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, \kappa(\boldsymbol{\vartheta}_l)\ . \tag{5.20}$$

Here, we distinguish by notation the sums over the positions $\boldsymbol{\vartheta}_{jkl}$ that are performed to calculate the aperture mass statistics at point $\boldsymbol{\phi}_i$ from the sum over all $\boldsymbol{\phi}_i$ that is performed to calculate the average of the aperture mass statistics.

5.2.1. The Gaussian aperture statistics covariance from the bispectrum covariance

As Eq. (2.133) describes how to model the third-order aperture statistics from the projected bispectrum B_κ, the straight-forward approach to derive the covariance of $\widehat{\mathcal{M}}_{\mathrm{ap}}^3$ is from the covariance of B_κ.

The bispectrum covariance has been derived Joachimi et al. (2009). In this section, we want to focus on the Gaussian part of the covariance, which is given by

$$\mathrm{C}_{B_\kappa}(\ell_1,\ell_2,\ell_3,\ell_4,\ell_5,\ell_6) = \frac{(2\pi)^6 \Lambda^{-1}(\ell_1,\ell_2,\ell_3)}{A\, A_R(\ell_1)\, A_R(\ell_2)\, A_R(\ell_3)}\ [\delta_{14}\,\delta_{25}\,\delta_{36} + 5 \text{ permutations}] \times P(\ell_1)\, P(\ell_2)\, P(\ell_3)\ , \tag{5.21}$$

where $A_R(\ell)$ is the size of the bin of $\boldsymbol{\ell}$, defined as

$$A_R(\ell) = 2\pi\, \ell\, \Delta\ell\ , \tag{5.22}$$

and Λ is defined as (see Joachimi et al., 2009)

$$\begin{aligned}\int \mathrm{d}^2\boldsymbol{\ell}_1 \int \mathrm{d}^2\boldsymbol{\ell}_2 \int \mathrm{d}^2\boldsymbol{\ell}_3 \;\delta_\mathrm{D}(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3) &= \int \mathrm{d}\ell_1 \int \mathrm{d}\ell_2 \int \mathrm{d}\ell_3 \;\ell_1\,\ell_2\,\ell_3 \\ &\quad\times \int \mathrm{d}\phi_1 \int \mathrm{d}\phi_2 \int \mathrm{d}\phi_3 \;\delta_\mathrm{D}(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3) \\ &\equiv \int \mathrm{d}\ell_1 \int \mathrm{d}\ell_2 \int \mathrm{d}\ell_3 \;\ell_1\,\ell_2\,\ell_3\,2\pi\,\Lambda(\ell_1,\ell_2,\ell_3)\;.\end{aligned} \tag{5.23}$$

We note that Eq. (5.21) was derived under the assumption that the Fourier transform of the survey window function is a Dirac-delta distribution, which is equivalent to the assumption that the survey area is infinite (or at least much larger than any relevant scales). We can now derive the covariance of $\widehat{\mathcal{M}}^3_\mathrm{ap}$ from C_B. Using Eq. (5.23), the third-order aperture statistics are

$$\begin{aligned}\left\langle\mathcal{M}^3_\mathrm{ap}\right\rangle(\theta_1,\theta_2,\theta_3) &= \int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\;\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3) \\ &\quad\times B_\kappa(\ell_1,\ell_2,\ell_3)\,(2\pi)^3\delta_\mathrm{D}(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3)\;. \\ &= \frac{1}{(2\pi)^6}\int\mathrm{d}\ell_1\int\mathrm{d}\ell_2\int\mathrm{d}\ell_3\;\ell_1\,\ell_2\,\ell_3\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3) \\ &\quad\times B_\kappa(\ell_1,\ell_2,\ell_3)\,(2\pi)^3\Lambda(\ell_1,\ell_2,\ell_3)\;.\end{aligned} \tag{5.24}$$

Replacing the ℓ-integrals by a discrete sum over the ℓ_i, we get

$$\begin{aligned}\left\langle\mathcal{M}^3_\mathrm{ap}\right\rangle(\theta_1,\theta_2,\theta_3) &= \frac{1}{(2\pi)^6}\sum_{ijk}\Delta\ell_i\,\Delta\ell_j\,\Delta\ell_k\,\ell_i\,\ell_j\,\ell_k\,\tilde{u}(\ell_i\,\theta_1)\,\tilde{u}(\ell_j\,\theta_2)\,\tilde{u}(\ell_k\,\theta_3) \\ &\quad\times B_\kappa(\ell_i,\ell_j,\ell_k)\,(2\pi)^3\Lambda(\ell_i,\ell_j,\ell_k)\;.\end{aligned} \tag{5.25}$$

Here, the $\Delta\ell_i$ are the bin sizes of the ℓ_i bins. Using Eq. (5.25), we derive $C_{\widehat{\mathcal{M}}^3_\mathrm{ap}}$ from C_{B_κ} with

$$\begin{aligned}C_{\widehat{\mathcal{M}}^3_\mathrm{ap}}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) &= \frac{1}{(2\pi)^6}\sum_{ijk}\sum_{lmn}\Delta\ell_i\,\Delta\ell_j\,\Delta\ell_k\,\Delta\ell_l\,\Delta\ell_m\,\Delta\ell_n\,\ell_i\,\ell_j\,\ell_k\,\ell_l\,\ell_m\,\ell_n \\ &\quad\times\tilde{u}(\ell_i\,\theta_1)\,\tilde{u}(\ell_j\,\theta_2)\,\tilde{u}(\ell_k\,\theta_3)\,\tilde{u}(\ell_l\,\theta_4)\,\tilde{u}(\ell_m\,\theta_5)\,\tilde{u}(\ell_n\,\theta_6) \\ &\quad\times C_{B_\kappa}(\ell_i,\ell_j,\ell_k,\ell_l,\ell_m,\ell_n)\,\Lambda(\ell_i,\ell_j,\ell_k)\,\Lambda(\ell_l,\ell_m,\ell_n)\;.\end{aligned} \tag{5.26}$$

So the Gaussian covariance of the aperture statistics is

$$
\begin{aligned}
&C_{\widehat{\mathcal{M}}^3_{\mathrm{ap}}}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad= \frac{1}{(2\pi)^6}\sum_{ijk}\sum_{lmn}\Delta\ell_i\,\Delta\ell_j\,\Delta\ell_k\,\Delta\ell_l\,\Delta\ell_m\,\Delta\ell_n\,\ell_i\,\ell_j\,\ell_k\,\ell_l\,\ell_m\,\ell_n\,\tilde{u}(\ell_i\,\theta_1)\,\tilde{u}(\ell_j\,\theta_2)\,\tilde{u}(\ell_k\,\theta_3) \\
&\qquad\times\tilde{u}(\ell_l\,\theta_4)\,\tilde{u}(\ell_m\,\theta_5)\,\tilde{u}(\ell_n\,\theta_6)\,\frac{(2\pi)^3\Lambda^{-1}(\ell_i,\ell_j,\ell_k)}{A\,\Delta\ell_i\,\Delta\ell_j\,\Delta\ell_k\,\ell_i\,\ell_j\,\ell_k}\,[\delta_{il}\,\delta_{jm}\,\delta_{kn}+5\text{ Perm.}] \\
&\qquad\times P(\ell_i)\,P(\ell_j)\,P(\ell_k)\,\Lambda(\ell_i,\ell_j,\ell_k)\,\Lambda(\ell_l,\ell_m,\ell_n) \\
&\quad= \frac{1}{(2\pi)^3\,A}\sum_{ijk}\sum_{lmn}\Delta\ell_l\,\Delta\ell_m\,\Delta\ell_n\,\ell_l\,\ell_m\,\ell_n\,\tilde{u}(\ell_i\,\theta_1)\,\tilde{u}(\ell_j\,\theta_2)\,\tilde{u}(\ell_k\,\theta_3) \\
&\qquad\times\tilde{u}(\ell_l\,\theta_4)\,\tilde{u}(\ell_m\,\theta_5)\,\tilde{u}(\ell_n\,\theta_6)\,[\delta_{il}\,\delta_{jm}\,\delta_{kn}+5\text{ Perm.}] \\
&\qquad\times P(\ell_i)\,P(\ell_j)\,P(\ell_k)\,\Lambda(\ell_l,\ell_m,\ell_n) \\
&\quad= \frac{1}{(2\pi)^3\,A}\sum_{ijk}\Delta\ell_i\,\Delta\ell_j\,\Delta\ell_k\,\ell_i\,\ell_j\,\ell_k\,\tilde{u}(\ell_i\,\theta_1)\,\tilde{u}(\ell_j\,\theta_2)\,\tilde{u}(\ell_k\,\theta_3) \\
&\qquad\times[\tilde{u}(\ell_l\,\theta_4)\,\tilde{u}(\ell_m\,\theta_5)\,\tilde{u}(\ell_n\,\theta_6)+5\text{ Perm.}]\;P(\ell_i)\,P(\ell_j)\,P(\ell_k)\,\Lambda(\ell_i,\ell_j,\ell_k)\;. \qquad (5.27)
\end{aligned}
$$

Assuming $\Delta\ell_i \to 0$, we can transform the sums back into integrals, yielding

$$
\begin{aligned}
C_{\widehat{\mathcal{M}}^3_{\mathrm{ap}}}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) &= \frac{1}{(2\pi)^3\,A}\int\mathrm{d}\ell_1\int\mathrm{d}\ell_2\int\mathrm{d}\ell_3\;\ell_1\,\ell_2\,\ell_3\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3) \\
&\quad\times[\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6)+5\text{ Perm.}] \\
&\quad\times P(\ell_1)\,P(\ell_2)\,P(\ell_3)\,\Lambda(\ell_1,\ell_2,\ell_3)\;.
\end{aligned}
$$

Transforming the Λ term from Eq. (5.23) back, we get

$$
\begin{aligned}
C_{\widehat{\mathcal{M}}^3_{\mathrm{ap}}}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) &= \frac{1}{(2\pi)^2\,A}\int\mathrm{d}^2\boldsymbol{\ell}_1\int\mathrm{d}^2\boldsymbol{\ell}_2\int\mathrm{d}^2\boldsymbol{\ell}_3\;\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3) \\
&\quad\times[\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6)+5\text{ Perm.}] \\
&\quad\times P(\ell_1)\,P(\ell_2)\,P(\ell_3)\,\delta_{\mathrm{D}}(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3) \\
&= \frac{1}{(2\pi)^3\,A}\int\mathrm{d}\ell_1\int\mathrm{d}\ell_2\int\mathrm{d}\phi\;\ell_1\,\ell_2\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3) \\
&\quad\times[\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6)+5\text{ Perm.}] \qquad (5.28) \\
&\quad\times P(\ell_1)\,P(\ell_2)\,P(\ell_3)\;,
\end{aligned}
$$

with $\ell_3^2=\ell_1^2+\ell_2^2+2\ell_1\,\ell_2\,\cos(\phi)$. This calculation can now easily be tested using Gaussian random fields: If we insert a known power spectrum, we know that the expectation value of $\widehat{\mathcal{M}}^3_{\mathrm{ap}}$ is zero, and the covariance $C_{\widehat{\mathcal{M}}^3_{\mathrm{ap}}}$ is purely Gaussian and should thus be given by Eq. (5.28). Inspecting Fig. 5.13, we see that this method can estimate the variance (i.e. the diagonal part of the covariance matrix) of third-order aperture statistics for most combinations of aperture radii with reasonable accuracy, but completely fails to predict the off-diagonal parts of the covariance matrix correctly. In the next section, we derive the covariance $C_{\widehat{\mathcal{M}}^3_{\mathrm{ap}}}$ directly from the estimator $\widehat{\mathcal{M}}^3_{\mathrm{ap}}$ and show that the method introduced here misses an important term due to the assumption that the Fourier transform of the survey window function is a Dirac-delta distribution.

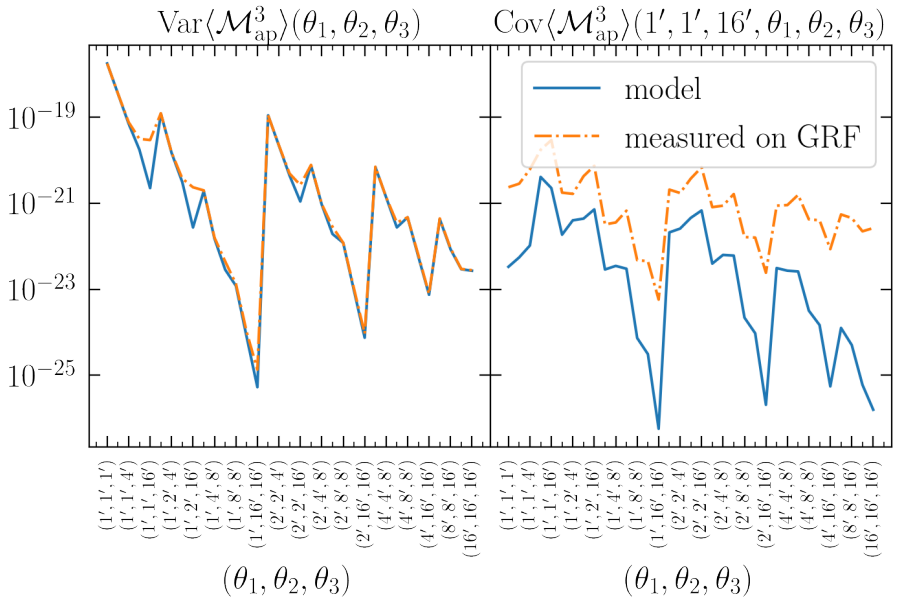


Figure 5.13. – We compare the model covariance from Eq. (5.28) with the one measured in Gaussian random fields (GRF). The left panel shows the Gaussian part variance of $\left\langle\mathcal{M}^3_{\mathrm{ap}}\right\rangle$ for a $10\times10\,\mathrm{deg}^2$ field with a constant power spectrum representing the case of a pure shape-noise signal. The x-axis iterates over all possible combinations of $(\theta_1,\theta_2,\theta_3)$ with $\theta_1\leq\theta_2\leq\theta_3$, where $\theta_i\in\{1',2',4',8',16'\}$. The right panel shows the covariance of the same $\left\langle\mathcal{M}^3_{\mathrm{ap}}\right\rangle$ statistics, where θ_1,θ_2 and θ_3 are fixed to the values $1',1'$ and $16'$, respectively.

5.2.2. The aperture statistics covariance from the direct estimator

Instead of deriving the aperture statistics covariance from the bispectrum covariance, we now directly derive the covariance for the estimator $\widehat{\mathcal{M}}^3_{\mathrm{ap}}$ of the aperture statistics $\left\langle\mathcal{M}^3_{\mathrm{ap}}\right\rangle$ from Eq. (5.20). The covariance of this estimator is defined as

$$\begin{aligned}C_{\widehat{\mathcal{M}}^3_{\mathrm{ap}}}&(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)\\&=\left\langle\widehat{\mathcal{M}}^3_{\mathrm{ap}}\,\widehat{\mathcal{M}}^3_{\mathrm{ap}}\right\rangle(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)-\left\langle\widehat{\mathcal{M}}^3_{\mathrm{ap}}(\theta_1,\theta_2,\theta_3)\right\rangle\left\langle\widehat{\mathcal{M}}^3_{\mathrm{ap}}(\theta_4,\theta_5,\theta_6)\right\rangle\,,\end{aligned}\tag{5.29}$$

with

$$\begin{aligned}&\left\langle\widehat{\mathcal{M}}^3_{\mathrm{ap}}\,\widehat{\mathcal{M}}^3_{\mathrm{ap}}\right\rangle(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)\\&\quad=\frac{A^6_{\mathrm{pix}}}{N^2_{\mathrm{pix}}}\sum_{i=1}^{N_{\mathrm{pix}}}\sum_{j=1}^{N_{\mathrm{pix}}}\sum_{klm}\sum_{npq}U_{\theta_1}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_k|)\,U_{\theta_2}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_l|)\,U_{\theta_3}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_m|)\\&\qquad\times U_{\theta_4}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_n|)\,U_{\theta_5}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_p|)\,U_{\theta_6}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_q|)\\&\qquad\times\left\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_m)\,\kappa(\boldsymbol{\vartheta}_n)\,\kappa(\boldsymbol{\vartheta}_p)\,\kappa(\boldsymbol{\vartheta}_q)\right\rangle\,.\end{aligned}\tag{5.30}$$

We want to write the six-point correlation function in terms of its connected correlation functions, denoted by $\langle \ldots \rangle_c$ (compare Sect. 2.1.2.4):

$$
\begin{aligned}
&\left\langle \widehat{\mathcal{M}}^3_{\rm ap}\, \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) \\
&\qquad = \frac{A^6_{\rm pix}}{N^2_{\rm pix}} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\qquad\quad \times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \\
&\qquad\quad \times \Big[\langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l) \rangle_c \, \langle \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_n) \rangle_c \, \langle \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_c + 14\,\text{Perm.} \\
&\qquad\qquad + \langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_m) \rangle_c \, \langle \kappa(\boldsymbol{\vartheta}_n)\, \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_c + 9\,\text{Perm.} \\
&\qquad\qquad + \langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l) \rangle_c \, \langle \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_n)\, \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_c + 14\,\text{Perm.} \\
&\qquad\qquad + \langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_n)\, \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_c \Big] \\
&\qquad = \left\langle \widehat{\mathcal{M}}^3_{\rm ap}\, \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle_{\rm G} (\theta_1, \ldots, \theta_6) + \left\langle \widehat{\mathcal{M}}^3_{\rm ap}\, \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle_{\rm NG,\,1} (\theta_1, \ldots, \theta_6) \\
&\qquad\quad + \left\langle \widehat{\mathcal{M}}^3_{\rm ap}\, \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle_{\rm NG,\,2} (\theta_1, \ldots, \theta_6) + \left\langle \widehat{\mathcal{M}}^3_{\rm ap}\, \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle_{\rm NG,\,3} (\theta_1, \ldots, \theta_6)
\end{aligned}
\tag{5.31}
$$

At first, we want to focus on the Gaussian part of the covariance to see if we can recover the covariance we measure in the GRF. Assuming that all connected correlations higher than second order vanish, what remains is the first term $\left\langle \widehat{\mathcal{M}}^3_{\rm ap}\, \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle_{\rm G}$, which depends only on the κ-2pcf. The other terms make up the *non-Gaussian* part of the covariance and will be assessed later.

5.2.2.1. The Gaussian part

The Gaussian part of the aperture mass covariance $\left\langle \widehat{\mathcal{M}}^3_{\rm ap} \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle_{\rm G}$ is given by

$$
\begin{aligned}
&\left\langle \widehat{\mathcal{M}}^3_{\rm ap} \widehat{\mathcal{M}}^3_{\rm ap} \right\rangle_{\rm G} (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) = \\
&\qquad \frac{A^6_{\rm pix}}{N^2_{\rm pix}} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\qquad\quad \times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \\
&\qquad\quad \times \langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l) \rangle_c \, \langle \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_n) \rangle_c \, \langle \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_c + 14\,\text{Perm.}
\end{aligned}
\tag{5.32}
$$

Inspecting Eq. (5.32), we can see that it is comprised of two groups, namely

$$\begin{aligned}
\left\langle \widehat{\mathcal{M}}_{\rm ap}^3 \widehat{\mathcal{M}}_{\rm ap}^3 \right\rangle_{\rm G}&(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) = \\
&\frac{A_{\rm pix}^6}{N_{\rm pix}^2} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\quad \times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \\
&\quad \times \left\langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_n) \right\rangle_c \left\langle \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_p) \right\rangle_c \left\langle \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_q) \right\rangle_c + 5\,\text{Perm.} \\
+ \quad &\frac{A_{\rm pix}^6}{N_{\rm pix}^2} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\quad \times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \\
&\quad \times \left\langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l) \right\rangle_c \left\langle \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_n) \right\rangle_c \left\langle \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \right\rangle_c + 8\,\text{Perm.} \\
&= T_1(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) + T_2(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)\,. \qquad (5.33)
\end{aligned}$$

The first group consists of six terms, in which for all three $\langle \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_m) \rangle_c$ the positions $\boldsymbol{\vartheta}_l, \boldsymbol{\vartheta}_m$ are subtracted from different positions $\boldsymbol{\phi}_i, \boldsymbol{\phi}_j$ in the U-filters. The second group consists of the nine other permutations, where in each case, one of the $\langle \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_m) \rangle_c$ has the same form as the first case, and for the other two both positions $\boldsymbol{\vartheta}_l, \boldsymbol{\vartheta}_m$ are subtracted from the same $\boldsymbol{\phi}_{i,j}$. To evaluate T_1 and T_2, we use Eq. (2.49) to replace the correlation functions by convergence power spectra and find

$$\begin{aligned}
T_1(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) = &\frac{A_{\rm pix}^6}{N_{\rm pix}^2} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \qquad (5.34) \\
&\times \int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\, P(\ell_1)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\ell}_1\cdot(\boldsymbol{\vartheta}_k - \boldsymbol{\vartheta}_n)} \int \frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\, P(\ell_2)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\ell}_2\cdot(\boldsymbol{\vartheta}_l - \boldsymbol{\vartheta}_p)} \\
&\times \int \frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\, P(\ell_3)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\ell}_3\cdot(\boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}_q)} + 5\ \text{Perm.}\,, \\
T_2(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) = &\frac{A_{\rm pix}^6}{N_{\rm pix}^2} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \qquad (5.35) \\
&\times \int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\, P(\ell_1)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\ell}_1\cdot(\boldsymbol{\vartheta}_k - \boldsymbol{\vartheta}_l)} \int \frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\, P(\ell_2)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\ell}_2\cdot(\boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}_n)} \\
&\times \int \frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\, P(\ell_3)\, \mathrm{e}^{-\mathrm{i}\boldsymbol{\ell}_3\cdot(\boldsymbol{\vartheta}_p - \boldsymbol{\vartheta}_q)} + 8\ \text{Perm.}\,.
\end{aligned}$$

We now replace the sums over (k,l,m,n,p,q) by integrals, assuming that the pixel size $A_{\rm pix}$ is small with respect to all relevant scales, and all relevant scales are small with respect to the survey area $A = N_{\rm pix}\, A_{\rm pix}$:

$$A_{\rm pix}^3 \sum_{klm} \to \int \mathrm{d}^2\vartheta_1 \int \mathrm{d}^2\vartheta_2 \int \mathrm{d}^2\vartheta_3\ . \qquad (5.36)$$

Under this assumption, we get

$$\begin{aligned} T_1(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) = \frac{1}{N_\mathrm{pix}^2}\sum_{i=1}^{N_\mathrm{pix}}\sum_{j=1}^{N_\mathrm{pix}}\int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\,P(\ell_1)\,P(\ell_2)\,P(\ell_3) \\ \times\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3)\,\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6) \\ \times\,\mathrm{e}^{-\mathrm{i}(\boldsymbol{\phi}_i-\boldsymbol{\phi}_j)\cdot(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3)} + 5\text{ Perm.}\,, \end{aligned} \tag{5.37}$$

$$\begin{aligned} T_2(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) = \frac{1}{N_\mathrm{pix}^2}\sum_{i=1}^{N_\mathrm{pix}}\sum_{j=1}^{N_\mathrm{pix}}\int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\,P(\ell_1)\,P(\ell_2)\,P(\ell_3) \\ \times\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_1\,\theta_2)\,\tilde{u}(\ell_2\,\theta_3)\,\tilde{u}(\ell_2\,\theta_4)\,\tilde{u}(\ell_3\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6) \\ \times\,\mathrm{e}^{-\mathrm{i}(\boldsymbol{\phi}_i-\boldsymbol{\phi}_j)\cdot\boldsymbol{\ell}_2} + 8\text{ Perm.}\,. \end{aligned} \tag{5.38}$$

Next, we replace the remaining sums with integrals. Here, we need to take into account the finite survey area A, and use

$$\frac{1}{N_\mathrm{pix}}\sum_{i=1}^{N_\mathrm{pix}} \to \frac{1}{A}\int_A \mathrm{d}^2\phi\ . \tag{5.39}$$

We further define the geometry factor $G_A(\boldsymbol{\ell})$ as

$$G_A(\boldsymbol{\ell}) = \frac{1}{A^2}\int_A \mathrm{d}^2\phi_1 \int_A \mathrm{d}^2\phi_2\ \mathrm{e}^{-\mathrm{i}(\boldsymbol{\phi}_1-\boldsymbol{\phi}_2)\cdot\boldsymbol{\ell}}\,. \tag{5.40}$$

We get

$$\begin{aligned} &T_1(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\ &\quad= \int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\,P(\ell_1)\,P(\ell_2)\,P(\ell_3)\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3) \\ &\qquad\times\,\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6)\,G_A(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2+\boldsymbol{\ell}_3) + 5\text{ Perm.}\,, \end{aligned} \tag{5.41}$$

$$\begin{aligned} &T_2(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\ &\quad= \int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\,P(\ell_1)\,P(\ell_2)\,P(\ell_3)\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_1\,\theta_2)\,\tilde{u}(\ell_2\,\theta_3) \\ &\qquad\times\,\tilde{u}(\ell_2\,\theta_4)\,\tilde{u}(\ell_3\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6)\,G_A(\boldsymbol{\ell}_2) + 8\text{ Perm.} \end{aligned} \tag{5.42}$$

For a square survey with side length L,

$$G_{A,\mathrm{square}}(\boldsymbol{\ell}) = \frac{2\sin^2(\ell_x\,L/2)}{\ell_x^2\,L^2}\,\frac{2\sin^2(\ell_y\,L/2)}{\ell_y^2\,L^2}\,, \tag{5.43}$$

where $\boldsymbol{\ell} = (\ell_x, \ell_y)$. For an (almost) infinite survey, however, we get (Joachimi et al., 2008)

$$G_A^\infty(\boldsymbol{\ell}) \to \frac{(2\pi)^2}{A}\delta_\mathrm{D}(\boldsymbol{\ell})\,, \tag{5.44}$$

In this case, we get $T_2^\infty = 0$, as $\tilde{u}(0) = 0$ holds for any compensated filter. Therefore, the large field approximation $C^\infty_{\widehat{\mathcal{M}^3_\mathrm{ap}}}$ of the Gaussian covariance is given solely by T_1^∞, which

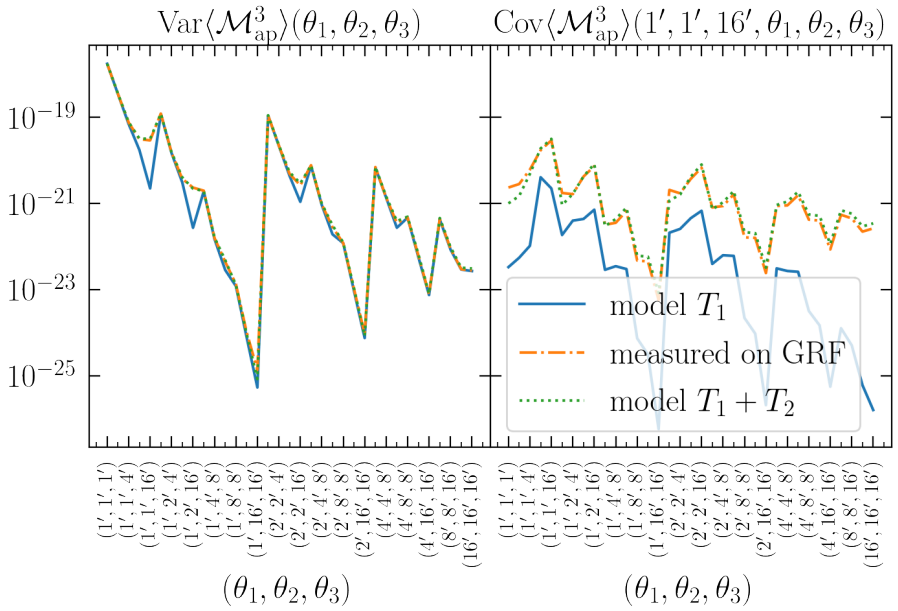


Figure 5.14. – Same as Fig. 5.13, but including the finite-field term T_2 in the model. The term T_1 corresponds to the model in Fig. 5.13.

becomes

$$
\begin{aligned}
&T_1^{\infty}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad= \frac{1}{(2\pi)^3\,A}\int_0^{\infty}\mathrm{d}\ell_1\int_0^{\infty}\mathrm{d}\ell_2\int_0^{2\pi}\mathrm{d}\phi\;\; P(\ell_1)\,P(\ell_2)\,P(\ell_3)\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(\ell_3\,\theta_3) \\
&\qquad\times\,\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_5)\,\tilde{u}(\ell_3\,\theta_6) + 5\text{ Perm.}\,,
\end{aligned}
\tag{5.45}
$$

where $\ell_3^2 = \ell_1^2 + \ell_2^2 + 2\ell_1\ell_2\cos(\phi)$. We can see that, in this case, we recover the Gaussian covariance from Eq. (5.28).

Inspecting Fig. 5.14, we can see that the finite field size indeed plays a significant role in the determination of the covariance for $\left\langle\mathcal{M}_{\mathrm{ap}}^3\right\rangle$, at least when small fields of $100\,\mathrm{deg}^2$ are considered. Conversely, this means that the covariance of third-order aperture statistics for a large survey can *not* be obtained by measuring it in a smaller simulation and then re-scaling it to fit the correct area. The difference is mainly driven by the term T_2, whereas the differences between T_1 and T_1^{∞} (which corresponds to Eq. 5.28) are subdominant.

5.2.2.2. The non-Gaussian part

We now want to derive the non-Gaussian part of the covariance using essentially the same steps as for the Gaussian part. As one can see in Eq. (5.31), these non-Gaussian parts depend on the three-, four- and six-point correlations of the convergence field. We thus introduce the bispectrum B, the trispectrum T and the pentaspectrum P_6 as their Fourier

counterparts (compare Eq. 2.52)

$$\begin{aligned} B(\ell_1,\ell_2,\ell_3) &= P_3(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)\;, \\ T(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3) &= P_4(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3,-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2-\boldsymbol{\ell}_3)\;, \\ P_6(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3,\boldsymbol{\ell}_4,\boldsymbol{\ell}_5) &= P_6(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3,\boldsymbol{\ell}_4,\boldsymbol{\ell}_5,-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2-\boldsymbol{\ell}_3-\boldsymbol{\ell}_4-\boldsymbol{\ell}_5)\;. \end{aligned} \tag{5.46}$$

The term $\left\langle \widehat{\mathcal{M}}^3_{\rm ap}\widehat{\mathcal{M}}^3_{\rm ap}\right\rangle_{\rm NG,\,1}$ can again be divided into two parts.

$$\begin{aligned} &\left\langle \widehat{\mathcal{M}}^3_{\rm ap}\widehat{\mathcal{M}}^3_{\rm ap}\right\rangle_{\rm NG,\,1}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\ &\quad= \frac{A^6_{\rm pix}}{N^2_{\rm pix}}\sum_{i=1}^{N_{\rm pix}}\sum_{j=1}^{N_{\rm pix}}\sum_{klm}\sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_k|)\,U_{\theta_2}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_l|)\,U_{\theta_3}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_m|) \\ &\qquad\times U_{\theta_4}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_n|)\,U_{\theta_5}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_p|)\,U_{\theta_6}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_q|) \\ &\qquad\times\big[\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_m)\rangle_{\rm c}\,\langle\kappa(\boldsymbol{\vartheta}_n)\,\kappa(\boldsymbol{\vartheta}_p)\,\kappa(\boldsymbol{\vartheta}_q)\rangle_{\rm c} \\ &\qquad\quad+\langle\kappa(\boldsymbol{\vartheta}_n)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_m)\rangle_{\rm c}\,\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_p)\,\kappa(\boldsymbol{\vartheta}_q)\rangle_{\rm c}+8\text{ Perm. }\big] \\ &\quad= T_3(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)+T_4(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)\;. \end{aligned} \tag{5.47}$$

In the first part, T_3, the $\boldsymbol{\vartheta}_k,\boldsymbol{\vartheta}_l,\boldsymbol{\vartheta}_m$ in the connected correlations $\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_m)\rangle_{\rm c}$ are all substracted from the same $\boldsymbol{\phi}_i$. In the second part, one of the $\boldsymbol{\vartheta}_k,\boldsymbol{\vartheta}_l,\boldsymbol{\vartheta}_m$ is substracted from the other $\boldsymbol{\phi}_j$. For T_3 we get

$$\begin{aligned} T_3(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) &= \frac{A^6_{\rm pix}}{N^2_{\rm pix}}\sum_{i=1}^{N_{\rm pix}}\sum_{j=1}^{N_{\rm pix}}\sum_{klm}\sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_k|)\,U_{\theta_2}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_l|)\,U_{\theta_3}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_m|) \\ &\qquad\times U_{\theta_4}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_n|)\,U_{\theta_5}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_p|)\,U_{\theta_6}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_q|) \\ &\qquad\times\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_m)\rangle_{\rm c}\,\langle\kappa(\boldsymbol{\vartheta}_n)\,\kappa(\boldsymbol{\vartheta}_p)\,\kappa(\boldsymbol{\vartheta}_q)\rangle_{\rm c} \\ &= \left\langle\widehat{\mathcal{M}}^3_{\rm ap}(\theta_1,\theta_2,\theta_3)\right\rangle\left\langle\widehat{\mathcal{M}}^3_{\rm ap}(\theta_4,\theta_5,\theta_6)\right\rangle\;, \end{aligned} \tag{5.48}$$

which cancels the averages of $\left\langle\mathcal{M}^3_{\rm ap}\right\rangle$ that are substracted in the definition of the covariance of the estimator in Eq. (5.29). The term T_4 can be calculated as

$$\begin{aligned} &T_4(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\ &\quad= \frac{A^6_{\rm pix}}{N^2_{\rm pix}}\sum_{i=1}^{N_{\rm pix}}\sum_{j=1}^{N_{\rm pix}}\sum_{klm}\sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_k|)\,U_{\theta_2}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_l|)\,U_{\theta_3}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_m|)\,U_{\theta_4}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_n|) \\ &\qquad\times U_{\theta_5}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_p|)\,U_{\theta_6}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_q|)\,\langle\kappa(\boldsymbol{\vartheta}_n)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_m)\rangle_{\rm c}\,\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_p)\,\kappa(\boldsymbol{\vartheta}_q)\rangle_{\rm c}+8\text{ Perm.} \\ &\quad= \int\frac{{\rm d}^2\ell_1}{(2\pi)^2}\int\frac{{\rm d}^2\ell_2}{(2\pi)^2}\int\frac{{\rm d}^2\ell_4}{(2\pi)^2}\int\frac{{\rm d}^2\ell_5}{(2\pi)^2}\,B_\kappa(\ell_1,\ell_2,|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|)\,B_\kappa(\ell_4,\ell_5,|\boldsymbol{\ell}_4+\boldsymbol{\ell}_5|) \\ &\qquad\times\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|\,\theta_3)\,\tilde{u}(\ell_4\,\theta_1)\,\tilde{u}(\ell_5\,\theta_5)\,\tilde{u}(|\boldsymbol{\ell}_4+\boldsymbol{\ell}_5|\,\theta_6)\,G_A(\boldsymbol{\ell}_1-\boldsymbol{\ell}_4)+8\text{ Perm.}\;, \end{aligned} \tag{5.49}$$

where we again replaced the sums with integrals and inserted the bispectrum.

Second, we calculate the covariance part $\left\langle \widehat{\mathcal{M}}_{\rm ap}^3 \widehat{\mathcal{M}}_{\rm ap}^3 \right\rangle_{\rm NG,\,2}$. We divide this term into two parts as

$$\begin{aligned}
&\left\langle \widehat{\mathcal{M}}_{\rm ap}^3 \widehat{\mathcal{M}}_{\rm ap}^3 \right\rangle_{\rm NG,\,2}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\qquad = \frac{A_{\rm pix}^6}{N_{\rm pix}^2} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\qquad\quad \times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \\
&\qquad\quad \times \big[\langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_n) \rangle_{\rm c} \, \langle \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_{\rm c} + 8 \text{ Perm.} \\
&\qquad\qquad + \langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l) \rangle_{\rm c} \, \langle \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_n)\, \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_{\rm c} + 5 \text{ Perm. } \big] \\
&\qquad = T_5(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) + T_6(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)\,. \qquad (5.50)
\end{aligned}$$

The first 9 permutations contain all terms, for which the $\boldsymbol{\vartheta}_l, \boldsymbol{\vartheta}_m$ in the two-point correlation $\langle \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_m) \rangle$ are subtracted from different $\boldsymbol{\phi}_i, \boldsymbol{\phi}_j$. The remaining six terms contain all other permutations. We again replace the sums with integrals and the correlation functions with polyspectra and find

$$\begin{aligned}
&T_5(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad = \int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_3}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_4}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_5}{(2\pi)^2}\, P(\ell_1)\, T(\boldsymbol{\ell}_3,\boldsymbol{\ell}_4,\boldsymbol{\ell}_5)\, \tilde{u}(\ell_1\,\theta_1)\, \tilde{u}(\ell_3\,\theta_2)\, \tilde{u}(\ell_4\,\theta_3) \\
&\qquad \times\, \tilde{u}(\ell_1\,\theta_4)\, \tilde{u}(\ell_5\,\theta_5)\, \tilde{u}(|\boldsymbol{\ell}_3 + \boldsymbol{\ell}_4 + \boldsymbol{\ell}_5|\,\theta_6)\, G_A(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_3 + \boldsymbol{\ell}_4) + 8 \text{ Perm.}\,, \qquad (5.51) \\
&T_6(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad = \int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_3}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_4}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_5}{(2\pi)^2}\, P(\ell_1)\, T(\boldsymbol{\ell}_3,\boldsymbol{\ell}_4,\boldsymbol{\ell}_5)\, \tilde{u}(\ell_1\,\theta_1)\, \tilde{u}(\ell_1\,\theta_2)\, \tilde{u}(\ell_3\,\theta_3) \\
&\qquad \times\, \tilde{u}(\ell_4\,\theta_4)\, \tilde{u}(\ell_5\,\theta_5)\, \tilde{u}(|\boldsymbol{\ell}_3 + \boldsymbol{\ell}_4 + \boldsymbol{\ell}_5|\,\theta_6)\, G_A(\boldsymbol{\ell}_3) + 5 \text{ Perm.}\,. \qquad (5.52)
\end{aligned}$$

At last, we consider $\left\langle \widehat{\mathcal{M}}_{\rm ap}^3 \widehat{\mathcal{M}}_{\rm ap}^3 \right\rangle_{\rm NG,\,3}$, which is

$$\begin{aligned}
&\left\langle \widehat{\mathcal{M}}_{\rm ap}^3 \widehat{\mathcal{M}}_{\rm ap}^3 \right\rangle_{\rm NG,\,3}(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\qquad = \frac{A_{\rm pix}^6}{N_{\rm pix}^2} \sum_{i=1}^{N_{\rm pix}} \sum_{j=1}^{N_{\rm pix}} \sum_{klm} \sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_k|)\, U_{\theta_2}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_l|)\, U_{\theta_3}(|\boldsymbol{\phi}_i - \boldsymbol{\vartheta}_m|) \\
&\qquad\quad \times U_{\theta_4}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_n|)\, U_{\theta_5}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_p|)\, U_{\theta_6}(|\boldsymbol{\phi}_j - \boldsymbol{\vartheta}_q|) \\
&\qquad\quad \times \langle \kappa(\boldsymbol{\vartheta}_k)\, \kappa(\boldsymbol{\vartheta}_l)\, \kappa(\boldsymbol{\vartheta}_m)\, \kappa(\boldsymbol{\vartheta}_n)\, \kappa(\boldsymbol{\vartheta}_p)\, \kappa(\boldsymbol{\vartheta}_q) \rangle_{\rm c} \\
&\qquad = T_7(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6)\,. \qquad (5.53)
\end{aligned}$$

Again, by replacing the sums with integrals, we get

$$\begin{aligned}
&T_7(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad = \int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_2}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_3}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_4}{(2\pi)^2} \int \frac{\mathrm{d}^2\ell_5}{(2\pi)^2}\, P_6(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,\boldsymbol{\ell}_3,\boldsymbol{\ell}_4,\boldsymbol{\ell}_5)\, \tilde{u}(\ell_1\,\theta_1)\, \tilde{u}(\ell_2\,\theta_2) \\
&\qquad \times\, \tilde{u}(\ell_3\,\theta_3)\, \tilde{u}(\ell_4\,\theta_4)\, \tilde{u}(\ell_5\,\theta_5)\, \tilde{u}(|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3 + \boldsymbol{\ell}_4 + \boldsymbol{\ell}_5|\,\theta_6)\, G_A(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3)\,. \qquad (5.54)
\end{aligned}$$

To test the behaviour for large survey areas, we can again replace the $G_A(\boldsymbol{\ell})$ in the terms T_4 to T_7 by Eq. (5.44) and find

$$
\begin{aligned}
&T_4^\infty(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad= \frac{1}{A}\int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_5}{(2\pi)^2}\,B_\kappa(\ell_1,\ell_2,|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|)\,B_\kappa(\ell_1,\ell_5,|\boldsymbol{\ell}_1+\boldsymbol{\ell}_5|) \qquad (5.55)\\
&\qquad\times\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|\,\theta_3)\,\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_5\,\theta_5)\,\tilde{u}(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_5|\,\theta_6)+8\text{ Perm.}\,,\\
&T_5^\infty(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad= \frac{1}{A,}\int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_3}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_5}{(2\pi)^2}\,P(\ell_1)\,T(\boldsymbol{\ell}_3,-\boldsymbol{\ell}_1-\boldsymbol{\ell}_3,\boldsymbol{\ell}_5) \qquad (5.56)\\
&\qquad\times\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_3\,\theta_2)\,\tilde{u}(|\boldsymbol{\ell}_1+\ell_3|\,\theta_3)\,\tilde{u}(\ell_1\,\theta_4)\,\tilde{u}(\ell_5\,\theta_5)\,\tilde{u}(|\boldsymbol{\ell}_5-\boldsymbol{\ell}_1|\,\theta_6)+8\text{ Perm.}\,,\\
&T_7^\infty(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6) \\
&\quad= \frac{1}{A,}\int\frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_2}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_4}{(2\pi)^2}\int\frac{\mathrm{d}^2\ell_5}{(2\pi)^2}\,P_6(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2,-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2,\boldsymbol{\ell}_4,\boldsymbol{\ell}_5) \qquad (5.57)\\
&\qquad\times\tilde{u}(\ell_1\,\theta_1)\,\tilde{u}(\ell_2\,\theta_2)\,\tilde{u}(|\boldsymbol{\ell}_1+\boldsymbol{\ell}_2|\,\theta_3)\,\tilde{u}(\ell_4\,\theta_4)\,\tilde{u}(\ell_5\,\theta_5)\,\tilde{u}(|\boldsymbol{\ell}_4+\boldsymbol{\ell}_5|\,\theta_6)\,.
\end{aligned}
$$

Like T_2^∞, the term T_6^∞ vanishes, as $\tilde{u}(0)=0$.

Unfortunately, we now face a problem: While we have well-tested and validated models for the power- and bispectrum, those are still lacking for the tri- and pentaspectrum. At this stage, we are satisfied with a relatively basic model of these spectra and thus choose to model them from their 1-halo terms within the halo model framework (see Sect. 2.1.2.6).

5.2.3. Measuring the covariance terms in simulations

Instead of validating the entire covariance, we have developed a method to validate the individual T_i-terms from the simulations. This is particularly interesting as it allows us to isolate the single contributions from different polyspectra to the final covariance. For example, one can extract the Gaussian covariance T_1+T_2 from a non-Gaussian field and see which covariance contributions dominate.

To explain our method, we want to focus on an arbitrary term of the full covariance in Eq. (5.31), for example, the term

$$
\begin{aligned}
&\frac{A_\mathrm{pix}^6}{N_\mathrm{pix}^2}\sum_{i=1}^{N_\mathrm{pix}}\sum_{j=1}^{N_\mathrm{pix}}\sum_{klm}\sum_{npq} U_{\theta_1}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_k|)\,U_{\theta_2}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_l|)\,U_{\theta_3}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_m|)\\
&\quad\times U_{\theta_4}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_n|)\,U_{\theta_5}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_p|)\,U_{\theta_6}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_q|)\\
&\quad\quad+\left\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_q)\right\rangle_c\left\langle\kappa(\boldsymbol{\vartheta}_n)\,\kappa(\boldsymbol{\vartheta}_p)\,\kappa(\boldsymbol{\vartheta}_m)\right\rangle_c\\
=\;&\frac{A_\mathrm{pix}^6}{N_\mathrm{pix}^2}\sum_{i=1}^{N_\mathrm{pix}}\sum_{j=1}^{N_\mathrm{pix}}\sum_{klq} U_{\theta_1}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_k|)\,U_{\theta_2}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_l|)\,U_{\theta_6}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_q|)\\
&\quad\times\left\langle\kappa(\boldsymbol{\vartheta}_k)\,\kappa(\boldsymbol{\vartheta}_l)\,\kappa(\boldsymbol{\vartheta}_q)\right\rangle_c\\
&\quad\times\sum_{npm} U_{\theta_4}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_n|)\,U_{\theta_5}(|\boldsymbol{\phi}_j-\boldsymbol{\vartheta}_p|)\,U_{\theta_3}(|\boldsymbol{\phi}_i-\boldsymbol{\vartheta}_m|)\\
&\quad\quad\times\left\langle\kappa(\boldsymbol{\vartheta}_n)\,\kappa(\boldsymbol{\vartheta}_p)\,\kappa(\boldsymbol{\vartheta}_m)\right\rangle_c\\
=\;&\frac{1}{N_\mathrm{pix}^2}\sum_{i=1}^{N_\mathrm{pix}}\sum_{j=1}^{N_\mathrm{pix}}\left\langle\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_i;\theta_1)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_i;\theta_2)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_j;\theta_6)\right\rangle_c\\
&\quad\times\left\langle\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_j;\theta_4)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_j;\theta_5)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_i;\theta_3)\right\rangle_c\\
\equiv\;&\frac{1}{N_\mathrm{pix}^2}\sum_{i=1}^{N_\mathrm{pix}}\sum_{j=1}^{N_\mathrm{pix}}\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_6;|\boldsymbol{\phi}_j-\boldsymbol{\phi}_i|)\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_4,\theta_5;\theta_3;|\boldsymbol{\phi}_i-\boldsymbol{\phi}_j|)\;. \qquad (5.58)
\end{aligned}
$$

Here, we have defined the quantity $\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_3;\eta)$, which is a two-point correlation function between the field $\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi};\theta_1)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi};\theta_2)$ and the field $\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi};\theta_3)$.

We can now again replace the sums over i and j by integrals, and replace the $\boldsymbol{\phi}_2$ integration by one over $\boldsymbol{\eta}=\boldsymbol{\phi}_2-\boldsymbol{\phi}_1$. Since the resulting integrand is independent of $\boldsymbol{\phi}_1$, we can perform the integration, which yields $AE(\boldsymbol{\eta})$, where the function $E(\boldsymbol{\eta})$ denotes the probability that for a random point $\boldsymbol{\phi}_1$ in A, the point $\boldsymbol{\phi}_1=\boldsymbol{\phi}_1+\boldsymbol{\eta}$ also lies in A (compare Heydenreich et al., 2020). We find for that term:

$$
\begin{aligned}
&\frac{1}{A^2}\int_A \mathrm{d}^2\phi_1\int_A \mathrm{d}^2\phi_2\;\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_6;|\boldsymbol{\phi}_j-\boldsymbol{\phi}_i|)\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_4,\theta_5;\theta_3;|\boldsymbol{\phi}_i-\boldsymbol{\phi}_j|)\\
=\;&\frac{1}{A}\int \mathrm{d}^2\eta\,E(\boldsymbol{\eta})\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_6;|\boldsymbol{\eta}|)\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_4,\theta_5;\theta_3;|\boldsymbol{\eta}|)\\
=\;&\frac{2\pi}{A}\int_0^\infty \mathrm{d}\eta\,\eta\,E(\eta)\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_6;\eta)\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_4,\theta_5;\theta_3;\eta)\;, \qquad (5.59)
\end{aligned}
$$

where in the last step we have defined $E(\eta)$ as the azimuthal average of $E(\boldsymbol{\eta})$. For a square survey of side-length L, Heydenreich et al. (2020) found

$$
E(\eta)=\begin{cases}\frac{1}{L^2\pi}\left[L^2\pi-(4L-\eta)\eta\right] & \eta\le L\\ \frac{2}{\pi}\left[4\sqrt{\frac{\eta^2}{L^2}-1}-1-\frac{\eta^2}{2L^2}-\operatorname{acos}\left(\frac{L}{\eta}\right)+\operatorname{asin}\left(\frac{L}{\eta}\right)\right] & L\le\eta\le\sqrt{2}L\\ 0 & \text{else}\end{cases}\;. \qquad (5.60)
$$

Inspecting Eq. (5.31) again, we can see that we can apply this method to every single one of the terms in this equation. In analogy to the definition of $\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_3;\eta)$, we define

$$\begin{aligned}
\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;|\boldsymbol{\phi}_2-\boldsymbol{\phi}_1|) &= \left\langle \widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_1)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2);\theta_2\right\rangle \\
\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_3;|\boldsymbol{\phi}_2-\boldsymbol{\phi}_1|) &= \left\langle \widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_1)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_2)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_3)\right\rangle \\
\xi^{4,1}_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2,\theta_3;\theta_4;|\boldsymbol{\phi}_2-\boldsymbol{\phi}_1|) &= \left\langle \widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_1)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_2)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_3)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_4)\right\rangle \\
\xi^{4,2}_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_3,\theta_4;|\boldsymbol{\phi}_2-\boldsymbol{\phi}_1|) &= \left\langle \widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_1)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_2)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_3)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_4)\right\rangle \\
\xi^6_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2,\theta_3;\theta_4,\theta_5,\theta_6;|\boldsymbol{\phi}_2-\boldsymbol{\phi}_1|) &= \Big\langle \widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_1)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_2)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_3) \\
&\quad \times\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_4)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_5)\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_6)\Big\rangle\ . \quad (5.61)
\end{aligned}$$

Applying the same calculations that lead to Eq. (5.59) to the terms T_1 and T_2 in Eq. (5.33), we see that every term of T_1 becomes a product of three correlation functions $\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;\eta)$, whereas in T_2 we always get a product of two second-order aperture masses $\left\langle\widehat{\mathcal{M}}^2_\mathrm{ap}(\theta_1,\theta_2)\right\rangle$ and one correlation function $\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;\eta)$. Similar observations can be made for all T_i terms defined in the previous section.

We note that in Eq. (5.61) we have switched from connected correlations $\langle\ldots\rangle_c$ to general correlations $\langle\ldots\rangle$. This makes no difference for second- and third-order correlations of a zero-mean field, but it does for the fourth- and sixth-order correlations we see in the other terms of Eq. (5.31). When dealing with these terms, we thus have to transform the non-connected $\widehat{T}_i^\mathrm{nc}$ into the connected $\widehat{T}_i$.

We can now calculate the T_i terms that can be estimated from simulations and get

$$\begin{aligned}
\widehat{T}_1(\theta_1,\ldots,\theta_6) &= \frac{2\pi}{A}\int_0^\infty \mathrm{d}\eta\,\eta E(\eta)\left[\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_4;\eta)\,\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_2;\theta_5;\eta)\,\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_3;\theta_6;\eta)+5\text{ Perm.}\right] \\
\widehat{T}_2(\theta_1,\ldots,\theta_6) &= \frac{2\pi}{A}\int_0^\infty \mathrm{d}\eta\,\eta E(\eta)\left[\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_4;\eta)\left\langle\widehat{\mathcal{M}}^2_\mathrm{ap}\right\rangle(\theta_2,\theta_3)\left\langle\widehat{\mathcal{M}}^2_\mathrm{ap}\right\rangle(\theta_5,\theta_6)+8\text{ Perm.}\right] \\
\widehat{T}_3(\theta_1,\ldots,\theta_6) &= \widehat{\mathcal{M}}^3_\mathrm{ap}(\theta_1,\theta_2,\theta_3)\widehat{\mathcal{M}}^3_\mathrm{ap}(\theta_4,\theta_5,\theta_6) \\
\widehat{T}_4(\theta_1,\ldots,\theta_6) &= \frac{2\pi}{A}\int_0^\infty \mathrm{d}\eta\,\eta E(\eta)\left[\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2;\theta_4;\eta)\,\xi^3_{\mathcal{M}_\mathrm{ap}}(\theta_5,\theta_6;\theta_3;\eta)+8\text{ Perm.}\right] \\
\widehat{T}^\mathrm{nc}_5(\theta_1,\ldots,\theta_6) &= \frac{2\pi}{A}\int_0^\infty \mathrm{d}\eta\,\eta E(\eta)\left[\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_4;\eta)\,\xi^{4,2}_{\mathcal{M}_\mathrm{ap}}(\theta_2,\theta_3;\theta_5,\theta_6;\eta)+8\text{ Perm.}\right] \\
\widehat{T}^\mathrm{nc}_6(\theta_1,\ldots,\theta_6) &= \frac{2\pi}{A}\int_0^\infty \mathrm{d}\eta\,\eta E(\eta)\left[\left\langle\widehat{\mathcal{M}}^2_\mathrm{ap}\right\rangle(\theta_1,\theta_2)\,\xi^{4,1}_{\mathcal{M}_\mathrm{ap}}(\theta_4,\theta_5,\theta_6;\theta_3;\eta)+5\text{ Perm.}\right] \\
\widehat{T}^\mathrm{nc}_7(\theta_1,\ldots,\theta_6) &= \frac{2\pi}{A}\int_0^\infty \mathrm{d}\eta\,\eta E(\eta)\,\xi^6_{\mathcal{M}_\mathrm{ap}}(\theta_1,\theta_2,\theta_3;\theta_4,\theta_5,\theta_6;\eta)\ . \quad (5.62)
\end{aligned}$$

To calculate the connected terms, we disentangle the fourth-order non-connected correlations into their components and verify that $\widehat{T}_5=\widehat{T}^\mathrm{nc}_5-3T_1-T_2$ and $\widehat{T}_6=\widehat{T}^\mathrm{nc}_6-2T_2$ holds. We further see that $\widehat{T}^\mathrm{nc}_7-\widehat{T}_3$ describes the full covariance as defined in Eq. (5.29), meaning $\widehat{T}_7=\widehat{T}^\mathrm{nc}_7-\widehat{T}_1-\widehat{T}_2-\widehat{T}_3-\widehat{T}_4-\widehat{T}_5-\widehat{T}_6$.

With this method, we can test the individual covariance terms that we calculated in Sect. 5.2.2.1 and 5.2.2.2. The computations of the correlation functions can even be

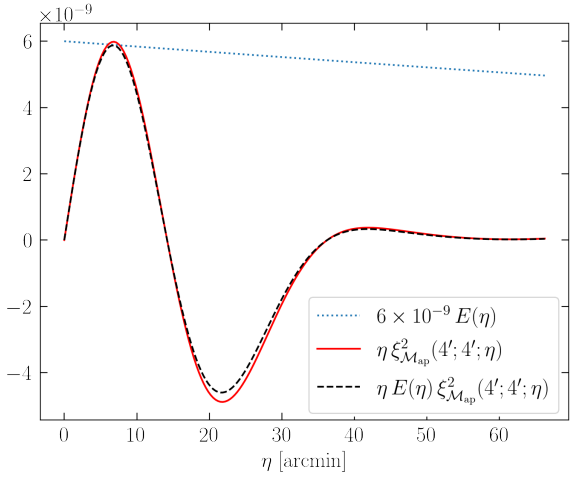


Figure 5.15. – The function $\eta\,\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;\eta)$ for $\theta_1 = \theta_2 = 4'$, measured in 640 convergence maps of the SLICS at $z \approx 0.8$. For comparison, we show the function $6 \times 10^{-9}\,E(\eta)$ for a survey of side-length $L = 472'$, which corresponds to the SLICS when $4 \times 16'$ are cut off from each side to avoid boundary effects. We also show the product of the two functions. We note that the integral over $\eta\,\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;\eta)$ is zero and the integral over the product of the two functions makes up the term $\widehat{T}_2$.

significantly sped up: For example, given fields $\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_1;\theta_1)$ and $\widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi}_2;\theta_2)$ defined on a square pixel grid, we can zero-pad these fields by adding pixel grid of the same size containing only zeros to the bottom, right, and bottom right of the original aperture mass field. When we then calculate the full correlation between them via a FFT, the result is precisely the correlation function multiplied with the function $E(\boldsymbol{\eta})$:

$$\int \mathrm{d}^2\phi\,\widehat{\mathcal{M}}_\mathrm{ap,padded}(\boldsymbol{\phi};\theta_1)\widehat{\mathcal{M}}_\mathrm{ap,padded}(\boldsymbol{\phi}+\boldsymbol{\eta};\theta_2) = E(\boldsymbol{\eta})\,\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;|\boldsymbol{\eta}|)\;. \tag{5.63}$$

Here, the fields $\widehat{\mathcal{M}}_\mathrm{ap,padded}(\boldsymbol{\phi};\theta)$ are defined as

$$\widehat{\mathcal{M}}_\mathrm{ap,padded}(\boldsymbol{\phi};\theta) = \begin{cases} \widehat{\mathcal{M}}_\mathrm{ap}(\boldsymbol{\phi};\theta) & \boldsymbol{\phi} \in A \\ 0 & \text{else} \end{cases}\;. \tag{5.64}$$

Inspecting Eq. (5.62), we notice that the term $\widehat{T}_2$ is quite unstable: The integral over the correlation function is compensated by definition, meaning

$$\int_0^\infty \mathrm{d}\eta\,\eta\,\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_4;\eta) = 0\;. \tag{5.65}$$

This means the only part that makes the integral for $\widehat{T}_2$ nonzero is multiplication with the slowly changing function $E(\eta)$. An example of this can be seen in Fig. 5.15. This is not entirely surprising, as the term T_2 vanishes for large fields. However, one can imagine that even a slight perturbation of the measured $\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;\eta)$ can lead to significant changes in the term $\widehat{T}_2$. A similar argument holds for the term $\widehat{T}_6$. This means that we have to measure the correlation functions $\xi^2_{\mathcal{M}_\mathrm{ap}}(\theta_1;\theta_2;\eta),\ldots$ as accurately as possible. To ensure

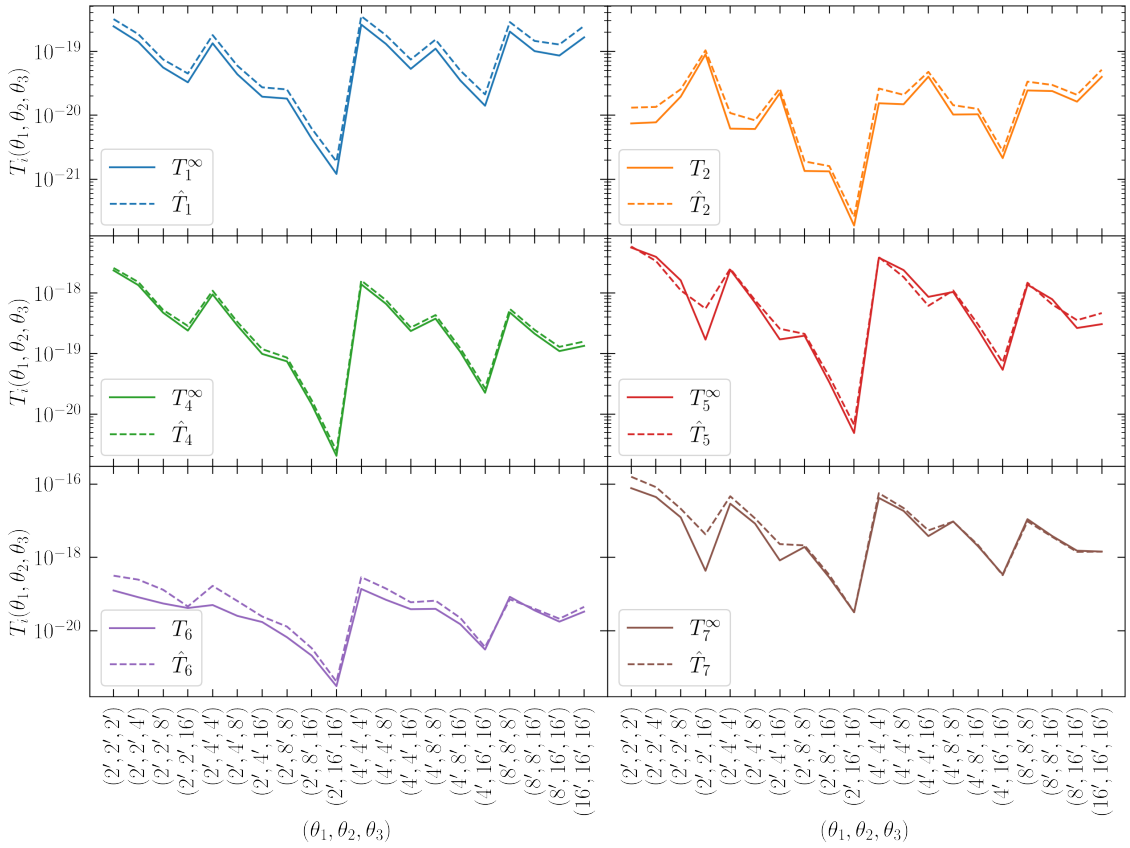


Figure 5.16. – The terms T_i calculated from the model (Sect. 5.2.2, solid lines) compared to the $\widehat{T}_i$ extracted from the 640 convergence maps of the SLICS at $z = 0.879$ (Sect. 5.2.3, dashed lines). The x axis represents all unique combinations of the aperture radii $2', 4', 8'$ and $16'$.

that, we take 640 noise-free lines of sight of convergence maps from the SLICS at redshift $z = 0.879$ to eliminate shape noise and minimise the sample variance as much as possible.

In Fig. 5.16, we compare the terms T_i that we modelled according to Sect. 5.2.2 with the $\widehat{T}_i$ described in Sect. 5.2.3. Overall, we notice a good agreement between the measurements and simulations. This means that we can model the individual terms of the covariance with reasonable accuracy. This is pleasantly surprising since our models for the tri- and pentaspectrum are relatively simple ones. We further see that the Gaussian part of the covariance, $T_1^\infty + T_2$, is completely subdominant to the non-Gaussian terms of the covariance; the term T_7 is by far the most dominant one. To test the resulting covariances, we compare the term $\widehat{T}_7^{\mathrm{mc}} - \widehat{T}_3$ with the sample covariance of the SLICS and the model covariance in Fig. 5.17. We see that, as expected, the two methods of measuring the covariance are extremely consistent. Between modelled and measured covariances, we notice a good agreement on larger scales but significant deviations on smaller scales. These deviations are particularly strong when combining small and large aperture radii.

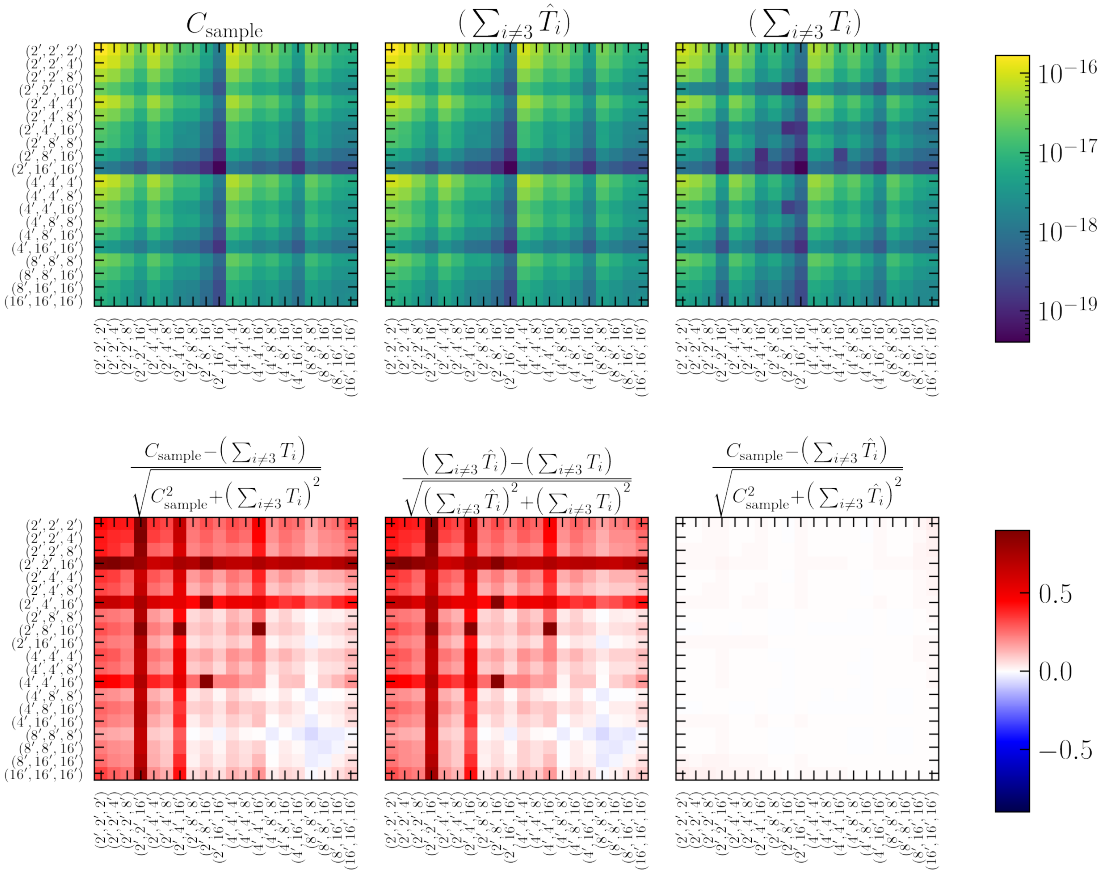


Figure 5.17. – The top row shows the sample covariance $C_{\rm sample}$, the covariance estimated from the measured $\widehat{T}_i$, $\left(\sum_{i\neq3}\widehat{T}_i\right) = \widehat{T}_7^{\rm nc} - \widehat{T}_3$, and the model covariance $\left(\sum_{i\neq3}T_i\right)$, respectively. The bottom row shows the relative differences between the covariance estimation methods.

5.3. Concluding remarks

In this section, we have laid the first stone to prepare a cosmological parameter analysis with third-order shear statistics. We have validated the BiHalofit bispectrum and concluded that the shear statistics derived from it are accurate enough for Stage-III surveys. Particularly we found that the flat-sky and Limber approximations are valid and that the model accuracy is mainly limited by the accuracy of the bispectrum model.

We have shown that third-order aperture statistics can be obtained in an unbiased way from the shear 3pcf if it is measured in an appropriate number of bins. For our analysis setup, 10^3 bins are sufficient for equal-scale aperture statistics, and 15^3 bins are a good choice if one wants to calculate the aperture statistics for all combinations of filter radii. We further verified and extended upon the results of Shi et al. (2014), who showed that the E- and B-mode leakage for third-order aperture statistics is at the per cent level.

We show that the information loss that arises when converting shear 3pcf to third-order aperture statistics is probably not severe and demonstrate that the computational load

of a parameter analysis with third-order aperture statistics is manageable, in particular when utilising emulators to predict the statistics.

In a mock analysis with both second- and third-order shear statistics and using a simulated covariance, we show that a combined analysis increases the constraining power on S_8 by almost a factor of two and further significantly improves the constraints on Ω_m and σ_8; the figure-of-merit increases by a factor of almost 6 and the individual constraints improve by 68% and 54%, respectively, mainly due to the different degeneracy directions between second- and third-order aperture statistics. While second-order aperture statistics are not used in cosmological parameter analyses, Asgari et al. (2021) showed that many second-order shear statistics exhibit similar constraining power and degeneracy directions, and we assume that second-order aperture statistics place no exception to that observation.

We find that equal-scale aperture statistics hold almost the same information content as the ones including all combinations of scales, which disagrees with the findings of Kilbinger and Schneider (2005) but is supported by Fu et al. (2014). This might suggest that a Fisher forecast is not an ideal tool to forecast cosmological parameter constraints, but further study is needed before this conclusion can be made. In particular, the impact of potential degeneracy breaking by the non-equal scale aperture mass statistics in higher-dimensional parameter spaces has not been investigated.

In future work, we plan to improve the measurement of the shear 3pcf. Philcox et al. (2022, and references therein) developed a method to estimate the n-point correlation function of galaxies in $\mathcal{O}(N^2)$ calculations by performing a spherical decomposition of the correlation function and then estimating the individual components. Although the $\mathcal{O}(N^2)$ scaling is worse than the $\mathcal{O}(N \log N)$ scaling of tree-codes, the total runtime appears to be faster in first tests[7]. Using a similar method, we can decompose the shear 3pcf into its multipole moments. Not only will this yield a faster measurement of the shear 3pcf, but we expect it to also improve the conversion to aperture mass statistics since the ψ-integration in Eq. (2.134) can be replaced by a sum over the multipole components.

We have further constructed a covariance model and shown that the naíve approach of simply transforming the bispectrum covariance yields wrong results. While the agreement between the simulated covariance and our model is not perfect, we are positively surprised by the level of agreement, considering we use a relatively simple model for the tri- and pentaspectrum. In the following months, we will work towards understanding the individual terms that contribute to the covariance, refine the covariance model and develop a joint covariance for second- and third-order shear statistics.

Before a cosmological parameter analysis can be performed, the final step is to understand the impacts of systematic biases on third-order shear statistics, in particular intrinsic alignments of source galaxies (Pyne et al., 2022) and baryonic feedback (Semboloni et al., 2008; Halder and Barreira, 2022).

[7] If one distributes the galaxies on a grid before the runtime can even be reduced to $\mathcal{O}(N \log N)$ when utilising the convolution theorem.

CHAPTER 6

Conclusion and outlook

In this thesis, we have worked towards establishing cosmological parameter analyses with higher-order shear statistics, both with a simulation-based inference method and with statistics that can be modelled directly. In both projects, we have shown that a combined analysis with second- and higher-order statistics can significantly increase the constraining power of a cosmic shear survey. On their own, summary statistics from persistent homology are significantly more constraining than third-order shear statistics. However, the improvement of the joint constraints in combination with second-order statistics appears to be better for third-order aperture masses. Furthermore, we have shown that higher-order statistics can detect peculiar phenomena that shear 2pcf are not sensitive to, as exemplified in the Ω_m-tension measured in Chap. 4.

As a simulation-based inference method, we have established persistent homology as an excellent higher-order statistic to constrain cosmological parameters. In this work, we were the first to quantify the constraining power of and perform a cosmological parameter analysis with persistent homology. In the first part of this project, we developed a cosmological infererence pipeline utilising persistent homology and showed that this method outperforms the commonly used peak count statistics by including information about the persistence of features, which is sensitive to the environment around overdensities in the matter distribution. We expect this information gain to become more critical in Stage-IV surveys. In the second part of the project, we improved our used summary statistics and extended the inference pipeline to include tomography and marginalisation over systematic effects. We then performed a cosmological parameter analysis of DES-Y1 data and compared the resulting constraints to the ones from shear 2pcf. We find $S_8 = 0.747^{+0.025}_{-0.031}$ for the matter clustering parameter, which is in full agreement with our analysis of second-order statistics ($S_8 = 0.759^{+0.049}_{-0.042}$) and other cosmological parameter analyses on the same data set (Troxel et al., 2018; Joudaki et al., 2020; Harnois-Déraps et al., 2021). We find a tension in the matter density parameter Ω_m, where we get $\Omega_\mathrm{m} = 0.468^{+0.051}_{-0.036}$ from persistent homology and $\Omega_\mathrm{m} = 0.256^{+0.034}_{-0.058}$ from shear 2pcf. This tension has been observed before by Harnois-Déraps et al. (2021) using peak count statistics. They used a com-

pletely independent analysis pipeline but utilised the same simulation set to model their data vector. If this tension is caused by a systematic bias in the inference pipeline, the underlying simulation set appears to be the most likely cause. However, none of the tests we performed indicated that this was the case. While we can not confidently exclude a statistical fluctuation or some systematic bias as the cause for this tension, the most exciting explanation would be some non-Gaussian features in the matter distribution that can not be captured by two-point statistics. To investigate whether this tension is due to a statistical fluctuation, we plan to perform a similar analysis on KiDS-1000 data.

In the second part of this thesis, we have focused on third-order aperture statistics as a higher-order statistic that can be modelled directly from the matter bispectrum. We have performed many validation tests of the convergence bispectrum, the shear 3pcf and the third-order aperture mass statistics, utilising different simulation suites and source redshift distributions. All these tests suggest that third-order shear statistics that are modelled from the BiHalofit bispectrum can be used to constrain cosmological parameters in Stage-III surveys. The biases introduced by the approximations of our shear inference pipeline (e.g. the reduced shear approximation, the Limber approximation, ...) are subdominant to the uncertainties in the bispectrum model. Until an improved bispectrum model is found, an analysis of a Stage-IV survey can not be performed. To prepare for a tomographic analysis, we have also developed and validated a model for the covariance of third-order aperture mass statistics. We then argue that third-order aperture statistics constitute an excellent summary statistic for a parameter analysis with third-order shear statistics: They can be modelled quickly and measured in a survey with a complex geometry without any bias by converting them from shear 3pcf, as long as the bins of the 3pcf are chosen appropriately. Furthermore, they allow a decomposition of the shear signal into E- and B-modes, yielding an additional null test. Lastly, their covariance can be directly modelled or quickly extracted from N-body simulations. Although the modelling procedure is both more complicated and numerically more demanding than a simulation-based inference, the large number of validation tests available to us will result in greater confidence in the achieved parameter constraints and provide immense value for further cosmological parameter analyses in cosmic shear.

For future work, we plan to develop a method that improves the measurement of the shear 3pcf and its conversion to third-order aperture statistics by measuring the multipole moments of the 3pcf. Furthermore, we plan to develop a model for the joint covariance between second- and third-order shear statistics. Lastly, we want to implement and validate models for mitigating systematic effects. These efforts will culminate in a cosmological parameter analysis of KiDS data, utilising both second- and third-order shear statistics.

Comparing simulation-based inference with direct modelling strategies, we can certainly say that both have their advantages and disadvantages. A simulation-based inference pipeline is incredibly versatile. Once the N-body simulations have been tailored to match a specific survey, one can quickly perform a parameter inference with any chosen summary statistic with the methods outlined in Chap. 4. For example, we were able to introduce a completely new method and use it to perform a cosmological parameter analysis on a current-generation survey within less than two years, and we could use the same pipeline to conduct an inference with any other summary statistic that can be quickly measured in simulations. To contrast this, we took three years to build and validate a modelling

pipeline for third-order shear statistics and their covariance, despite a significant amount of previous efforts (Scoccimarro and Couchman, 2001; Schneider and Lombardi, 2003; Jarvis et al., 2004; Schneider et al., 2005; Kilbinger and Schneider, 2005; Semboloni et al., 2008; Joachimi et al., 2009; Gil-Marín et al., 2012; Fu et al., 2014; Takahashi et al., 2020, and many more), and we have not even included tomography or systematic effects. However, the great advantage of this method is that we are able to better understand our summary statistics. While the bispectrum is calibrated using N-body simulations, its general shape is based on a physically motivated model. All other parts of the modelling pipeline can be tested in an isolated setting, so we can exactly tell under which conditions they are accurate. In a simulation-based inference, where the model is simply interpolated from N-body simulations, it is harder to exclude the possibility that some simulation artefacts have become part of the model, meaning that one can not place as much trust in the final parameter constraints.

Within the next decade, we hope that our efforts will contribute to achieving tighter and more robust constraints on cosmological parameters so that we can either alleviate or confirm the tension in the parameter S_8. In the latter case, we plan to contribute to the testing of new models that improve upon the current ΛCDM model.

Bibliography

Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. (2018). The Dark Energy Survey: Data Release 1. *ApJS*, 239(2):18.

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. (2022). Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 105(2):023520.

Abdalla, E., Abellán, G. F., Aboubrahim, A., et al. (2022). Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies. *Journal of High Energy Astrophysics*, 34:49–211.

Aihara, H., Arimoto, N., Armstrong, R., et al. (2018). The Hyper Suprime-Cam SSP Survey: Overview and survey design. *PASJ*, 70:S4.

Albrecht, A., Bernstein, G., Cahn, R., et al. (2006). Report of the Dark Energy Task Force. *arXiv e-prints*, astro–ph/0609591.

Alcock, C., Allsman, R. A., Alves, D. R., et al. (2000). The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations. *ApJ*, 542(1):281–307.

Alonso, D., Sanchez, J., Slosar, A., and LSST Dark Energy Science Collaboration (2019). A unified pseudo-C_ℓ framework. *MNRAS*, 484(3):4127–4151.

Amon, A., Gruen, D., Troxel, M. A., et al. (2022). Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to data calibration. *Phys. Rev. D*, 105(2):023514.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley.

Angulo, R. E., Zennaro, M., Contreras, S., et al. (2021). The BACCO simulation project: exploiting the full power of large-scale structure for cosmology. *MNRAS*, 507(4):5869–5881.

Applegate, D. E., Mantz, A., Allen, S. W., et al. (2016). Cosmology and astrophysics from relaxed galaxy clusters - IV. Robustly calibrating hydrostatic masses with weak lensing. *MNRAS*, 457(2):1522–1534.

Asgari, M. and Heymans, C. (2019). Dark Energy Survey Year 1: An independent E/B-mode cosmic shear analysis. *MNRAS*, 484(1):L59–L63.

Asgari, M., Heymans, C., Hildebrandt, H., et al. (2019). Consistent cosmic shear in the face of systematics: a B-mode analysis of KiDS-450, DES-SV and CFHTLenS. *A&A*, 624:A134.

Asgari, M., Lin, C.-A., Joachimi, B., et al. (2021). KiDS-1000 cosmology: Cosmic shear constraints and comparison between two point statistics. *A&A*, 645:A104.

Asgari, M. and Schneider, P. (2015). A new data compression method and its application to cosmic shear analysis. *A&A*, 578:A50.

Asgari, M., Tröster, T., Heymans, C., et al. (2020). KiDS+VIKING-450 and DES-Y1 combined: Mitigating baryon feedback uncertainty with COSEBIs. *A&A*, 634:A127.

Bartelmann, M. (2010). TOPICAL REVIEW Gravitational lensing. *Classical and Quantum Gravity*, 27(23):233001.

Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. *Phys. Rep.*, 340(4-5):291–472.

Beck, D., Fabbian, G., and Errard, J. (2018). Lensing reconstruction in post-Born cosmic microwave background weak lensing. *Phys. Rev. D*, 98(4):043512.

Benítez, N. (2000). Bayesian Photometric Redshift Estimation. *ApJ*, 536(2):571–583.

Bernardeau, F., van Waerbeke, L., and Mellier, Y. (1997). Weak lensing statistics as a probe of {OMEGA} and power spectrum. *A&A*, 322:1–18.

Berntsen, J., Espelid, T. O., and Genz, A. (1991). An adaptive algorithm for the approximate calculation of multiple integrals. *j-TOMS*, 17(4):437–451.

Biagetti, M., Calles, J., Castiblanco, L., Cole, A., and Noreña, J. (2022). Fisher Forecasts for Primordial non-Gaussianity from Persistent Homology. *arXiv e-prints*, arXiv:2203.08262.

Biagetti, M., Cole, A., and Shiu, G. (2020). The Persistence of Large Scale Structures I: Primordial non-Gaussianity. *arXiv e-prints*, arXiv:2009.04819.

BICEP2 Collaboration, Ade, P. A. R., Aikin, R. W., et al. (2014). Detection of B-Mode Polarization at Degree Angular Scales by BICEP2. *Phys. Rev. Lett.*, 112(24):241101.

Blakeslee, J. P., Jensen, J. B., Ma, C.-P., Milne, P. A., and Greene, J. E. (2021). The Hubble Constant from Infrared Surface Brightness Fluctuation Distances. *ApJ*, 911(1):65.

Blazek, J. A., MacCrann, N., Troxel, M. A., and Fang, X. (2019). Beyond linear galaxy alignments. *Phys. Rev. D*, 100(10):103506.

Bridle, S. and King, L. (2007). Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements. *New Journal of Physics*, 9(12):444.

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research (JMLR)*, 16:77–102.

Burger, P., Friedrich, O., Harnois-Déraps, J., and Schneider, P. (2022). A revised density split statistic model for general filters. *A&A*, 661:A137.

Burger, P., Schneider, P., Demchenko, V., et al. (2020). An adapted filter function for density split statistics in weak lensing. *A&A*, 642:A161.

Castro, T., Borgani, S., Dolag, K., et al. (2021). On the impact of baryons on the halo mass function, bias, and cluster cosmology. *MNRAS*, 500(2):2316–2335.

Castro, T., Quartin, M., Giocoli, C., Borgani, S., and Dolag, K. (2018). The effect of baryons in the cosmological lensing PDFs. *MNRAS*, 478(1):1305–1325.

Cea, P. (2022). CMB two-point angular correlation function in the Ellipsoidal Universe. *arXiv e-prints*, arXiv:2203.14229.

Chazal, F., de Silva, V., Glisse, M., and Oudot, S. (2016). *The Structure and Stability of Persistence Modules*. Springer International Publishing.

Chazal, F. and Michel, B. (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, 4:108.

Cheng, S., Ting, Y.-S., Ménard, B., and Bruna, J. (2020). A new approach to observational cosmology using the scattering transform. *MNRAS*, 499(4):5902–5914.

Chisari, N. E., Mead, A. J., Joudaki, S., et al. (2019). Modelling baryonic feedback for survey cosmology. *The Open Journal of Astrophysics*, 2(1):4.

Cooray, A. and Hu, W. (2002). Second-Order Corrections to Weak Lensing by Large-Scale Structure. *ApJ*, 574(1):19–23.

Crittenden, R. G., Natarajan, P., Pen, U.-L., and Theuns, T. (2002). Discriminating Weak Lensing from Intrinsic Spin Correlations Using the Curl-Gradient Decomposition. *ApJ*, 568(1):20–27.

Cuesta-Lazaro, C., Quera-Bofarull, A., Reischke, R., and Schäfer, B. M. (2018). Gravitational corrections to light propagation in a perturbed FLRW universe and corresponding weak-lensing spectra. *MNRAS*, 477(1):741–754.

Davis, P., Rabinowitz, P., and Rheinbolt, W. (2014). *Methods of Numerical Integration*. Computer Science and Applied Mathematics. Elsevier Science.

Davis, T. M. and Lineweaver, C. H. (2004). Expanding Confusion: Common Misconceptions of Cosmological Horizons and the Superluminal Expansion of the Universe. *PASA*, 21(1):97–109.

de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. (2015). The first and second data releases of the Kilo-Degree Survey. *A&A*, 582:A62.

de Jong, J. T. A., Verdoes Kleijn, G. A., Erben, T., et al. (2017). The third data release of the Kilo-Degree Survey and associated data products. *A&A*, 604:A134.

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., and Valentijn, E. A. (2013). The Kilo-Degree Survey. *Experimental Astronomy*, 35(1-2):25–44.

Deshpande, A. C. and Kitching, T. D. (2020). Post-Limber weak lensing bispectrum, reduced shear correction, and magnification bias correction. *Phys. Rev. D*, 101(10):103531.

Deshpande, A. C., Kitching, T. D., Cardone, V. F., et al. (2020a). Euclid: The reduced shear approximation and magnification bias for Stage IV cosmic shear experiments. *A&A*, 636:A95.

Deshpande, A. C., Taylor, P. L., and Kitching, T. D. (2020b). Accessing the high-ℓ frontier under the reduced shear approximation with k -cut cosmic shear. *Phys. Rev. D*, 102(8):083535.

Di Valentino, E., Anchordoqui, L. A., Akarsu, Ö., et al. (2021a). Cosmology Intertwined II: The hubble constant tension. *Astroparticle Physics*, 131:102605.

Di Valentino, E., Anchordoqui, L. A., Akarsu, Ö., et al. (2021b). Cosmology Intertwined III: fσ_8 and S_8. *Astroparticle Physics*, 131:102604.

Di Valentino, E., Anchordoqui, L. A., Akarsu, Ö., et al. (2021c). Cosmology Intertwined IV: The age of the universe and its curvature. *Astroparticle Physics*, 131:102607.

Dicke, R. (1970). *Gravitation and the Universe*. American Philosophical Society: Memoirs of the American Philosophical Society. American Philosophical Society.

Dicke, R. H., Peebles, P. J. E., Roll, P. G., and Wilkinson, D. T. (1965). Cosmic Black-Body Radiation. *ApJ*, 142:414–419.

Dodelson, S. and Schmidt, F. (2020). *Modern Cosmology*. Elsevier Science.

Dodelson, S. and Zhang, P. (2005). Weak lensing bispectrum. *Phys. Rev. D*, 72(8):083001.

Dubois, Y., Pichon, C., Welker, C., et al. (2014). Dancing in the dark: galactic properties trace spin swings along the cosmic web. *MNRAS*, 444(2):1453–1468.

Dyson, F. W., Eddington, A. S., and Davidson, C. (1920). A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London Series A*, 220:291–333.

Edge, A., Sutherland, W., Kuijken, K., et al. (2013). The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift. *The Messenger*, 154:32–34.

Einstein, A. (1915). Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin*, 844–847.

Eisenstein, D. J. and Hu, W. (1999). Power Spectra for Cold Dark Matter and Its Variants. *ApJ*, 511(1):5–15.

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. (2011). SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *AJ*, 142(3):72.

Elbers, W. and van de Weygaert, R. (2019). Persistent topology of the reionization bubble network - I. Formalism and phenomenology. *MNRAS*, 486(2):1523–1538.

Ellis, G. F. R. (2006). Issues in the Philosophy of Cosmology. *arXiv e-prints*, astro–ph/0602280.

Ellis, G. F. R., Maartens, R., and MacCallum, M. A. H. (2012). *Relativistic Cosmology*.

Elsner, F., Leistedt, B., and Peiris, H. V. (2017). Unbiased pseudo-C_ℓ power spectrum estimation with mode projection. *MNRAS*, 465(2):1847–1855.

Etherington, I. M. H. (1933). On the Definition of Distance in General Relativity. *Philosophical Magazine*, 15(18):761.

Falco, E. E., Gorenstein, M. V., and Shapiro, I. I. (1985). On model-dependent bounds on H 0 from gravitational images : application to Q 0957+561 A, B. *ApJ*, 289:L1–L4.

Flaugher, B., Diehl, H. T., Honscheid, K., et al. (2015). The Dark Energy Camera. *AJ*, 150(5):150.

Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386.

Fu, L., Kilbinger, M., Erben, T., et al. (2014). CFHTLenS: cosmological constraints from a combination of cosmic shear two-point and three-point correlations. *MNRAS*, 441(3):2725–2743.

Garnavich, P. M., Jha, S., Challis, P., et al. (1998). Supernova Limits on the Cosmic Equation of State. *ApJ*, 509(1):74–79.

Geis, N., Grupp, F., Prieto, E., and Bender, R. (2015). Preliminary results on the EUCLID NISP stray-light and ghost analysis. In MacEwen, H. A. and Breckinridge, J. B., editors, *Optics for EUV, X-Ray, and Gamma-Ray Astronomy VII*, volume 9602 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 96020H.

Genz, A. C. and Malik, A. A. (1980). Remarks on algorithm 006: An adaptive algorithm for numerical integration over an N-dimensional rectangular region. *Journal of Computational and Applied Mathematics*, 6(4):295–302.

Giblin, B., Heymans, C., Asgari, M., et al. (2021). KiDS-1000 catalogue: Weak gravitational lensing shear measurements. *A&A*, 645:A105.

Giblin, B., Heymans, C., Harnois-Déraps, J., et al. (2018). KiDS-450: enhancing cosmic shear with clipping transformations. *MNRAS*, 480(4):5529–5549.

Giblin, J. T., Mertens, J. B., Starkman, G. D., and Zentner, A. R. (2017). General relativistic corrections to the weak lensing convergence power spectrum. *Phys. Rev. D*, 96(10):103530.

Gil-Marín, H., Wagner, C., Fragkoudi, F., Jimenez, R., and Verde, L. (2012). An improved fitting formula for the dark matter bispectrum. *J. Cosmology Astropart. Phys.*, 2012(2):047.

Gorenstein, M. V., Falco, E. E., and Shapiro, I. I. (1988). Degeneracies in Parameter Estimates for Models of Gravitational Lens Systems. *ApJ*, 327:693.

Grimm, N. and Yoo, J. (2018). Jacobi mapping approach for a precise cosmological weak lensing formalism. *J. Cosmology Astropart. Phys.*, 2018(7):067.

Grimm, N. and Yoo, J. (2021). General relativistic effects in weak lensing angular power spectra. *Phys. Rev. D*, 104(8):083548.

Gruen, D. and Brimioulle, F. (2017). Selection biases in empirical p(z) methods for weak lensing. *MNRAS*, 468(1):769–782.

Gruen, D., Friedrich, O., Krause, E., et al. (2018). Density split statistics: Cosmological constraints from counts and lensing in cells in DES Y1 and SDSS data. *Phys. Rev. D*, 98(2):023507.

Gruen, D., Seitz, S., Koppenhoefer, J., and Riffeser, A. (2010). Bias-free Shear Estimation Using Artificial Neural Networks. *ApJ*, 720(1):639–651.

Guth, A. H., Kaiser, D. I., and Nomura, Y. (2014). Inflationary paradigm after Planck 2013. *Physics Letters B*, 733:112–119.

Guyonnet, A., Astier, P., Antilogus, P., Regnault, N., and Doherty, P. (2015). Evidence for self-interaction of charge distribution in charge-coupled devices. *A&A*, 575:A41.

Halder, A. and Barreira, A. (2022). Response approach to the integrated shear 3-point correlation function: the impact of baryonic effects on small scales. *MNRAS*.

Halder, A., Friedrich, O., Seitz, S., and Varga, T. N. (2021). The integrated three-point correlation function of cosmic shear. *MNRAS*, 506(2):2780–2803.

Hamana, T., Shirasaki, M., Miyazaki, S., et al. (2020). Cosmological constraints from cosmic shear two-point correlation functions with HSC survey first-year data. *PASJ*, 72(1):16.

Hamilton, A. J. S. (2000). Uncorrelated modes of the non-linear power spectrum. *MNRAS*, 312(2):257–284.

Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015). POLYCHORD: next-generation nested sampling. *MNRAS*, 453(4):4384–4398.

Harnois-Déraps, J., Amon, A., Choi, A., et al. (2018). Cosmological simulations for combined-probe analyses: covariance and neighbour-exclusion bias. *MNRAS*, 481(1):1337–1367.

Harnois-Déraps, J., Giblin, B., and Joachimi, B. (2019). Cosmic shear covariance matrix in wCDM: Cosmology matters. *A&A*, 631:A160.

Harnois-Déraps, J., Martinet, N., Castro, T., et al. (2021). Cosmic shear cosmology beyond two-point statistics: a combined peak count and correlation function analysis of DES-Y1. *MNRAS*, 506(2):1623–1650.

Hartlap, J., Hilbert, S., Schneider, P., and Hildebrandt, H. (2011). A bias in cosmic shear from galaxy selection: results from ray-tracing simulations. *A&A*, 528:A51.

Hartlap, J., Simon, P., and Schneider, P. (2007). Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *A&A*, 464(1):399–404.

Hatcher, A. (2002). *Algebraic Topology*. Algebraic Topology. Cambridge University Press.

Heydenreich, S., Brück, B., Burger, P., et al. (2022a). Persistent homology in cosmic shear II: A tomographic analysis of DES-Y1. *arXiv e-prints*, arXiv:2204.11831.

Heydenreich, S., Brück, B., and Harnois-Déraps, J. (2021). Persistent homology in cosmic shear: Constraining parameters with topological data analysis. *A&A*, 648:A74.

Heydenreich, S., Linke, L., Burger, P., and Schneider, P. (2022b). A roadmap to cosmological parameter analysis with third-order shear statistics I: Modelling and validation. *arXiv e-prints*, arXiv:2208.11686.

Heydenreich, S., Schneider, P., Hildebrandt, H., et al. (2020). The effects of varying depth in cosmic shear surveys. *A&A*, 634:A104.

Heymans, C., Grocutt, E., Heavens, A., et al. (2013). CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments. *MNRAS*, 432(3):2433–2453.

Heymans, C., Tröster, T., Asgari, M., et al. (2021). KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. *A&A*, 646:A140.

Hikage, C., Oguri, M., Hamana, T., et al. (2019). Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data. *PASJ*, 71(2):43.

Hilbert, S., Barreira, A., Fabbian, G., et al. (2020). The accuracy of weak lensing simulations. *MNRAS*, 493(1):305–319.

Hilbert, S., Hartlap, J., and Schneider, P. (2011). Cosmic shear covariance: the log-normal approximation. *A&A*, 536:A85.

Hilbert, S., Hartlap, J., White, S. D. M., and Schneider, P. (2009). Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. *A&A*, 499(1):31–43.

Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. (2020). KiDS+VIKING-450: Cosmic shear tomography with optical and infrared data. *A&A*, 633:A69.

Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. (2021). KiDS-1000 catalogue: Redshift distributions and their calibration. *A&A*, 647:A124.

Hildebrandt, H., Viola, M., Heymans, C., et al. (2017). KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *MNRAS*, 465(2):1454–1498.

Hirschmann, M., Dolag, K., Saro, A., et al. (2014). Cosmological simulations of black hole growth: AGN luminosities and downsizing. *MNRAS*, 442(3):2304–2324.

Hivon, E., Górski, K. M., Netterfield, C. B., et al. (2002). MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets. *ApJ*, 567(1):2–17.

Hoekstra, H., Franx, M., Kuijken, K., and Squires, G. (1998). Weak Lensing Analysis of CL 1358+62 Using Hubble Space Telescope Observations. *ApJ*, 504(2):636–660.

Hong, S. E., Park, S., Jee, M. J., Bak, D., and Cha, S. (2021). Weak-lensing Mass Reconstruction of Galaxy Clusters with a Convolutional Neural Network. *ApJ*, 923(2):266.

Hoyle, B., Gruen, D., Bernstein, G. M., et al. (2018). Dark Energy Survey Year 1 Results: redshift distributions of the weak-lensing source galaxies. *MNRAS*, 478(1):592–610.

Huang, H.-J., Eifler, T., Mandelbaum, R., and Dodelson, S. (2019). Modelling baryonic physics in future weak lensing surveys. *MNRAS*, 488(2):1652–1678.

Hubble, E. (1929a). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173.

Hubble, E. P. (1929b). A spiral nebula as a stellar system, Messier 31. *ApJ*, 69:103–158.

Huff, E. and Mandelbaum, R. (2017). Metacalibration: Direct Self-Calibration of Biases in Shear Measurement. *arXiv e-prints*, arXiv:1702.02600.

Ivezic, Z., Axelrod, T., Brandt, W. N., et al. (2008). Large Synoptic Survey Telescope: From Science Drivers To Reference Design. *Serbian Astronomical Journal*, 176:1–13.

James, O., Tunzelmann, E. v., Franklin, P., and Thorne, K. S. (2015). Gravitational lensing by spinning black holes in astrophysics, and in the movie Interstellar. *Classical and Quantum Gravity*, 32(6):065001.

Jarvis, M., Bernstein, G., and Jain, B. (2004). The skewness of the aperture mass statistic. *MNRAS*, 352(1):338–352.

Jee, M. J., Tyson, J. A., Schneider, M. D., et al. (2013). Cosmic Shear Results from the Deep Lens Survey. I. Joint Constraints on Ω_M and σ_8 with a Two-dimensional Analysis. *ApJ*, 765(1):74.

Jeffrey, N., Alsing, J., and Lanusse, F. (2021). Likelihood-free inference with neural compression of DES SV weak lensing map statistics. *MNRAS*, 501(1):954–969.

Joachimi, B., Lin, C. A., Asgari, M., et al. (2021). KiDS-1000 methodology: Modelling and inference for joint weak gravitational lensing and spectroscopic galaxy clustering analysis. *A&A*, 646:A129.

Joachimi, B., Schneider, P., and Eifler, T. (2008). Analysis of two-point statistics of cosmic shear. III. Covariances of shear measures made easy. *A&A*, 477(1):43–54.

Joachimi, B., Shi, X., and Schneider, P. (2009). Bispectrum covariance in the flat-sky limit. *A&A*, 508(3):1193–1204.

Joudaki, S., Blake, C., Heymans, C., et al. (2017). CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics. *MNRAS*, 465(2):2033–2052.

Joudaki, S., Hildebrandt, H., Traykova, D., et al. (2020). KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear. *A&A*, 638:L1.

Kacprzak, T., Kirk, D., Friedrich, O., et al. (2016). Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data. *MNRAS*, 463(4):3653–3673.

Kaiser, N. (1992). Weak Gravitational Lensing of Distant Galaxies. *ApJ*, 388:272.

Kaiser, N. (1998). Weak Lensing and Cosmology. *ApJ*, 498(1):26–42.

Kaiser, N. and Squires, G. (1993). Mapping the Dark Matter with Weak Gravitational Lensing. *ApJ*, 404:441.

Kaiser, N., Squires, G., and Broadhurst, T. (1995). A Method for Weak Lensing Observations. *ApJ*, 449:460.

Kannawadi, A., Shapiro, C. A., Mandelbaum, R., et al. (2016). The Impact of Interpixel Capacitance in CMOS Detectors on PSF Shapes and Implications for WFIRST. *PASP*, 128(967):095001.

Kayo, I., Takada, M., and Jain, B. (2013). Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix. *MNRAS*, 429(1):344–371.

Kilbinger, M., Fu, L., Heymans, C., et al. (2013). CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing. *MNRAS*, 430(3):2200–2220.

Kilbinger, M., Heymans, C., Asgari, M., et al. (2017). Precision calculations of the cosmic shear power spectrum projection. *MNRAS*, 472(2):2126–2141.

Kilbinger, M. and Schneider, P. (2005). Cosmological parameters from combined second- and third-order aperture mass statistics of cosmic shear. *A&A*, 442(1):69–83.

Kilbinger, M., Schneider, P., and Eifler, T. (2006). E- and B-mode mixing from incomplete knowledge of the shear correlation. *A&A*, 457(1):15–19.

Kuijken, K., Heymans, C., Dvornik, A., et al. (2019). The fourth data release of the Kilo-Degree Survey: ugri imaging and nine-band optical-IR photometry over 1000 square degrees. *A&A*, 625:A2.

Kuijken, K., Heymans, C., Hildebrandt, H., et al. (2015). Gravitational lensing analysis of the Kilo-Degree Survey. *MNRAS*, 454(4):3500–3532.

Laureijs, R., Amiaux, J., Arduini, S., et al. (2011). Euclid Definition Study Report. *arXiv e-prints*, arXiv:1110.3193.

Lemaître, G. (1931). Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *MNRAS*, 91:483–490.

Lemos, P., Weaverdyck, N., Rollins, R. P., et al. (2022). Robust sampling for weak lensing and clustering analyses with the Dark Energy Survey. *arXiv e-prints*, arXiv:2202.08233.

Lepage, G. P. (1978). A new algorithm for adaptive multidimensional integration. *Journal of Computational Physics*, 27(2):192–203.

Levin, D. (1982). Procedures for computing one- and two-dimensional integrals of functions with rapid irregular oscillations. *Mathematics of Computation*, 38(158):531–538.

Levin, D. (1996). Fast integration of rapidly oscillatory functions. *Journal of Computational and Applied Mathematics*, 67(1):95–101.

Lewis, A. and Challinor, A. (2011). CAMB: Code for Anisotropies in the Microwave Background. Astrophysics Source Code Library, record ascl:1102.026.

Lima, M., Cunha, C. E., Oyaizu, H., et al. (2008). Estimating the redshift distribution of photometric galaxy samples. *MNRAS*, 390(1):118–130.

Limber, D. N. (1954). The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II. *ApJ*, 119:655.

Linke, L., Simon, P., Schneider, P., and Hilbert, S. (2020). Measuring galaxy-galaxy-galaxy-lensing with higher precision and accuracy. *A&A*, 634:A13.

Liu, J., Petri, A., Haiman, Z., et al. (2015a). Cosmology constraints from the weak lensing peak counts and the power spectrum in CFHTLenS data. *Phys. Rev. D*, 91(6):063507.

Liu, X., Pan, C., Li, R., et al. (2015b). Cosmological constraints from weak lensing peak statistics with Canada-France-Hawaii Telescope Stripe 82 Survey. *MNRAS*, 450(3):2888–2902.

Lo, D. and Park, B. (2018). Modeling the spread of the Zika virus using topological data analysis. *PLoS ONE*, 13(2):e0192120.

Luongo, O., Muccino, M., Colgáin, E. Ó., Sheikh-Jabbari, M. M., and Yin, L. (2022). Larger H_0 values in the CMB dipole direction. *Phys. Rev. D*, 105(10):103510.

Luppino, G. A. and Kaiser, N. (1997). Detection of Weak Lensing by a Cluster of Galaxies at z = 0.83. *ApJ*, 475(1):20–28.

Mackey, J., White, M., and Kamionkowski, M. (2002). Theoretical estimates of intrinsic galaxy alignment. *MNRAS*, 332(4):788–798.

Makarenko, I., Shukurov, A., Henderson, R., et al. (2018). Topological signatures of interstellar magnetic fields - I. Betti numbers and persistence diagrams. *MNRAS*, 475(2):1843–1858.

Marques, G. A., Liu, J., Zorrilla Matilla, J. M., et al. (2019). Constraining neutrino mass with weak lensing Minkowski Functionals. *J. Cosmology Astropart. Phys.*, 2019(6):019.

Martinet, N., Harnois-Déraps, J., Jullo, E., and Schneider, P. (2021). Probing dark energy with tomographic weak-lensing aperture mass statistics. *A&A*, 646:A62.

Martinet, N., Schneider, P., Hildebrandt, H., et al. (2018). KiDS-450: cosmological constraints from weak-lensing peak statistics - II: Inference from shear peaks using N-body simulations. *MNRAS*, 474(1):712–730.

McCarthy, I. G., Schaye, J., Bird, S., and Le Brun, A. M. C. (2017). The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology. *MNRAS*, 465(3):2936–2965.

Migkas, K., Pacaud, F., Schellenberger, G., et al. (2021). Cosmological implications of the anisotropy of ten galaxy cluster scaling relations. *A&A*, 649:A151.

Migkas, K., Schellenberger, G., Reiprich, T. H., et al. (2020). Probing cosmic isotropy with a new X-ray galaxy cluster sample through the L_X-T scaling relation. *A&A*, 636:A15.

Migliaccio, M. and Planck Collaboration (2019). Cosmic microwave background cosmology with Planck. *Nuovo Cimento C Geophysics Space Physics C*, 41(4):132.

Miller, L. and CFHTLenS Collaboration (2012). Weak Lensing Shape Measurement in CFHTLenS. In *American Astronomical Society Meeting Abstracts #219*, volume 219 of *American Astronomical Society Meeting Abstracts*, 130.03.

Minami, Y. and Komatsu, E. (2020). New Extraction of the Cosmic Birefringence from the Planck 2018 Polarization Data. *Phys. Rev. Lett.*, 125(22):221301.

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1996). The Structure of Cold Dark Matter Halos. *ApJ*, 462:563.

Nicolau, M., Levine, A. J., and Carlsson, G. (2011). Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Science*, 108(17):7265–7270.

Ogata, H. (2005). A numerical integration formula based on the bessel functions. *Publications of the Research Institute for Mathematical Sciences*, 41:949–970.

Oudot, S. Y. (2015). *Persistence theory: from quiver representations to data analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Parroni, C., Cardone, V. F., Maoli, R., and Scaramella, R. (2020). Going deep with Minkowski functionals of convergence maps. *A&A*, 633:A71.

Peacock, J. A., Schneider, P., Efstathiou, G., et al. (2006). ESA-ESO Working Group on "Fundamental Cosmology". "ESA-ESO Working Group on "Fundamental Cosmology", Edited by J.A. Peacock et al. ESA, 2006.".

Peebles, P. (1980). *The Large-scale Structure of the Universe*. Princeton Series in Physics. Princeton University Press.

Peebles, P. (2020). *Principles of Physical Cosmology*. Princeton Series in Physics. Princeton University Press.

Peebles, P. J. E. (1966). Primordial Helium Abundance and the Primordial Fireball. II. *ApJ*, 146:542.

Peebles, P. J. E. (1982). Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations. *ApJ*, 263:L1–L5.

Penrose, R. (2016). *The Road to Reality: A Complete Guide to the Laws of the Universe*. Random House.

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421.

Percival, W. J., Friedrich, O., Sellentin, E., and Heavens, A. (2022). Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix. *MNRAS*, 510(3):3207–3221.

Pesce, D. W., Braatz, J. A., Reid, M. J., et al. (2020). The Megamaser Cosmology Project. XIII. Combined Hubble Constant Constraints. *ApJ*, 891(1):L1.

Petri, A., Haiman, Z., and May, M. (2017). Validity of the Born approximation for beyond Gaussian weak lensing observables. *Phys. Rev. D*, 95(12):123503.

Petri, A., Liu, J., Haiman, Z., et al. (2015). Emulating the CFHTLenS weak lensing data: Cosmological constraints from moments and Minkowski functionals. *Phys. Rev. D*, 91(10):103511.

Philcox, O. H. E. (2022). Probing Parity-Violation with the Four-Point Correlation Function of BOSS Galaxies. *arXiv e-prints*, arXiv:2206.04227.

Philcox, O. H. E., Slepian, Z., Hou, J., et al. (2022). ENCORE: an O ($N_g{}^2$) estimator for galaxy N-point correlation functions. *MNRAS*, 509(2):2457–2481.

Pillepich, A., Springel, V., Nelson, D., et al. (2018). Simulating galaxy formation with the IllustrisTNG model. *MNRAS*, 473(3):4077–4106.

Planck Collaboration, Aghanim, N., Akrami, Y., et al. (2020a). Planck 2018 results. I. Overview and the cosmological legacy of Planck. *A&A*, 641:A1.

Planck Collaboration, Aghanim, N., Akrami, Y., et al. (2020b). Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6.

Plazas, A. A., Shapiro, C., Kannawadi, A., et al. (2016). The Effect of Detector Nonlinearity on WFIRST PSF Profiles for Weak Gravitational Lensing Measurements. *PASP*, 128(968):104001.

Pontoppidan, K., Blome, C., Braun, H., et al. (2022). The JWST Early Release Observations. *arXiv e-prints*, arXiv:2207.13067.

Porqueres, N., Heavens, A., Mortlock, D., and Lavaux, G. (2021). Bayesian forward modelling of cosmic shear data. *MNRAS*, 502(2):3035–3044.

Porqueres, N., Heavens, A., Mortlock, D., and Lavaux, G. (2022). Lifting weak lensing degeneracies with a field-based likelihood. *MNRAS*, 509(3):3194–3202.

Pranav, P. (2021). Topology and geometry of Gaussian random fields II: on critical points, excursion sets, and persistent homology. *arXiv e-prints*, arXiv:2109.08721.

Pranav, P., Adler, R. J., Buchert, T., et al. (2019). Unexpected topology of the temperature fluctuations in the cosmic microwave background. *A&A*, 627:A163.

Pranav, P., Edelsbrunner, H., van de Weygaert, R., et al. (2017). The topology of the cosmic web in terms of persistent Betti numbers. *MNRAS*, 465(4):4281–4310.

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438.

Pyne, S. and Joachimi, B. (2021). Self-calibration of weak lensing systematic effects using combined two- and three-point statistics. *MNRAS*, 503(2):2300–2317.

Pyne, S., Tenneti, A., and Joachimi, B. (2022). Three-point intrinsic alignments of dark matter halos in the IllustrisTNG simulation. *arXiv e-prints*, arXiv:2204.10342.

Rieke, M. J., Kelly, D., and Horner, S. (2005). Overview of James Webb Space Telescope and NIRCam's Role. In Heaney, J. B. and Burriesci, L. G., editors, *Cryogenic Optical Systems and Instruments XI*, volume 5904 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 1–8.

Riess, A. G., Yuan, W., Macri, L. M., et al. (2022). A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km s^{-1} Mpc^{-1} Uncertainty from the Hubble Space Telescope and the SH0ES Team. *ApJ*, 934(1):L7.

Rindler, W. (1956). Visual horizons in world models. *MNRAS*, 116:662.

Rizzato, M. and Sellentin, E. (2022). Extremely expensive likelihoods: A variational-Bayes solution for precision cosmology. *arXiv e-prints*, arXiv:2203.05009.

Robertson, H. P. (1935). Kinematics and World-Structure. *ApJ*, 82:284.

Rosenbluth, M. N. (2003). Genesis of the Monte Carlo Algorithm for Statistical Mechanics. In Gubernatis, J. E., editor, *The Monte Carlo Method in the Physical Sciences*, volume 690 of *American Institute of Physics Conference Series*, 22–30.

Rubin, V. C. and Ford, W. Kent, J. (1970). Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379.

Sato, M. and Nishimichi, T. (2013). Impact of the non-Gaussian covariance of the weak lensing power spectrum and bispectrum on cosmological parameter estimation. *Phys. Rev. D*, 87(12):123538.

Schirmer, M., Erben, T., Hetterscheidt, M., and Schneider, P. (2007). GaBoDS: the Garching-Bonn Deep Survey. IX. A sample of 158 shear-selected mass concentration candidates. *A&A*, 462(3):875–887.

Schneider, P. (2003). The consequences of parity symmetry for higher-order statistics of cosmic shear and other polar fields. *A&A*, 408:829–834.

Schneider, P. (2005). Weak Gravitational Lensing. *arXiv e-prints*, astro–ph/0509252.

Schneider, P. (2006). *Extragalactic Astronomy and Cosmology*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Schneider, P. (2009). *Cosmology – Lecture Notes*.

Schneider, P. (2012). *Gravitational lensing – Lecture Notes*.

Schneider, P., Asgari, M., Jozani, Y. N., et al. (2022). Pure-mode correlation functions for cosmic shear and application to KiDS-1000. *A&A*, 664:A77.

Schneider, P., Eifler, T., and Krause, E. (2010). COSEBIs: Extracting the full E-/B-mode information from cosmic shear correlation functions. *A&A*, 520:A116.

Schneider, P., Kilbinger, M., and Lombardi, M. (2005). The three-point correlation function of cosmic shear. II. Relation to the bispectrum of the projected mass density and generalized third-order aperture measures. *A&A*, 431:9–25.

Schneider, P. and Lombardi, M. (2003). The three-point correlation function of cosmic shear. I. The natural components. *A&A*, 397:809–818.

Schneider, P. and Seitz, C. (1995). Steps towards nonlinear cluster inversion through gravitational distortions. I. Basic considerations and circular clusters. *A&A*, 294:411–431.

Schneider, P. and Sluse, D. (2013). Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing. *A&A*, 559:A37.

Schneider, P., van Waerbeke, L., Jain, B., and Kruse, G. (1998). A new measure for cosmic shear. *MNRAS*, 296(4):873–892.

Schneider, P., van Waerbeke, L., and Mellier, Y. (2002). B-modes in cosmic shear from source redshift clustering. *A&A*, 389:729–741.

Schrabback, T., Applegate, D., Dietrich, J. P., et al. (2018). Cluster mass calibration at high redshift: HST weak lensing analysis of 13 distant galaxy clusters from the South Pole Telescope Sunyaev-Zel'dovich Survey. *MNRAS*, 474(2):2635–2678.

Schrabback, T., Bocquet, S., Sommer, M., et al. (2021). Mass calibration of distant SPT galaxy clusters through expanded weak-lensing follow-up observations with HST, VLT, & Gemini-South. *MNRAS*, 505(3):3923–3943.

Scoccimarro, R. and Couchman, H. M. P. (2001). A fitting formula for the non-linear evolution of the bispectrum. *MNRAS*, 325(4):1312–1316.

Scoccimarro, R., Sheth, R. K., Hui, L., and Jain, B. (2001). How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *ApJ*, 546(1):20–34.

Secco, L. F., Jarvis, M., Jain, B., et al. (2022a). Dark Energy Survey Year 3 Results: Three-point shear correlations and mass aperture moments. *Phys. Rev. D*, 105(10):103537.

Secco, L. F., Samuroff, S., Krause, E., et al. (2022b). Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to modeling uncertainty. *Phys. Rev. D*, 105(2):023515.

Seitz, S. and Schneider, P. (2001). A new finite-field mass reconstruction algorithm. *A&A*, 374:740–745.

Seitz, S., Schneider, P., and Ehlers, J. (1994). Light propagation in arbitrary spacetimes and the gravitational lens approximation. *Classical and Quantum Gravity*, 11(9):2345–2373.

Sellentin, E. and Heavens, A. F. (2016). Parameter inference with estimated covariance matrices. *MNRAS*, 456(1):L132–L136.

Sellentin, E. and Heavens, A. F. (2018). On the insufficiency of arbitrarily precise covariance matrices: non-Gaussian weak-lensing likelihoods. *MNRAS*, 473(2):2355–2363.

Sellentin, E., Heymans, C., and Harnois-Déraps, J. (2018). The skewed weak lensing likelihood: why biases arise, despite data and theory being sound. *MNRAS*, 477(4):4879–4895.

Semboloni, E., Heymans, C., van Waerbeke, L., and Schneider, P. (2008). Sources of contamination to weak lensing three-point statistics: constraints from N-body simulations. *MNRAS*, 388(3):991–1000.

Shapiro, C. and Cooray, A. (2006). The Born and lens lens corrections to weak gravitational lensing angular power spectra. *J. Cosmology Astropart. Phys.*, 2006(3):007.

Sheldon, E. S., Becker, M. R., MacCrann, N., and Jarvis, M. (2020). Mitigating Shear-dependent Object Detection Biases with Metacalibration. *ApJ*, 902(2):138.

Sheldon, E. S. and Huff, E. M. (2017). Practical Weak-lensing Shear Measurement with Metacalibration. *ApJ*, 841(1):24.

Sheth, R. K. and Tormen, G. (1999). Large-scale bias and the peak background split. *MNRAS*, 308(1):119–126.

Shi, X., Joachimi, B., and Schneider, P. (2014). How well do third-order aperture mass statistics separate E- and B-modes? *A&A*, 561:A68.

Shi, X., Schneider, P., and Joachimi, B. (2011). Relations between three-point configuration space shear and convergence statistics. *A&A*, 533:A48.

Solà Peracaula, J., Gómez-Valent, A., and de Cruz Pérez, J. (2019). Signs of dynamical dark energy in current observations. *Physics of the Dark Universe*, 25:100311.

Sousbie, T. (2011). The persistent cosmic web and its filamentary structure - I. Theory and implementation. *MNRAS*, 414(1):350–383.

Springel, V., Pakmor, R., Zier, O., and Reinecke, M. (2021). Simulating cosmic structure formation with the GADGET-4 code. *MNRAS*, 506(2):2871–2949.

Springel, V., Yoshida, N., and White, S. D. M. (2001). GADGET: a code for collisionless and gasdynamical cosmological simulations. *New A*, 6(2):79–117.

Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., and Hobson, M. P. (2022). COSMOPOWER: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys. *MNRAS*, 511(2):1771–1788.

Steinhardt, P. J. (2011). The Inflation Debate. *Scientific American*, 304(4):36–43.

Stuart, A. and Ord, K. (2009). *Kendall's Advanced Theory of Statistics: Volume 1: Distribution Theory*. Number Bd. 1;Bd. 1994 in Kendall's Advanced Theory of Statistics. Wiley.

Takada, M. and Hu, W. (2013). Power spectrum super-sample covariance. *Phys. Rev. D*, 87(12):123504.

Takahashi, R., Hamana, T., Shirasaki, M., et al. (2017). Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments. *ApJ*, 850(1):24.

Takahashi, R., Nishimichi, T., Namikawa, T., et al. (2020). Fitting the Nonlinear Matter Bispectrum by the Halofit Approach. *ApJ*, 895(2):113.

Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., and Oguri, M. (2012). Revising the Halofit Model for the Nonlinear Matter Power Spectrum. *ApJ*, 761(2):152.

Talman, J. D. (1978). Numerical Fourier and Bessel Transforms in Logarithmic Variables. *Journal of Computational Physics*, 29(1):35–48.

Teklu, A. F., Remus, R.-S., Dolag, K., et al. (2015). Connecting Angular Momentum and Galactic Dynamics: The Complex Interplay between Spin, Mass, and Morphology. *ApJ*, 812(1):29.

Troxel, M. A. and Ishak, M. (2015). The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. *Phys. Rep.*, 558:1–59.

Troxel, M. A., MacCrann, N., Zuntz, J., et al. (2018). Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear. *Phys. Rev. D*, 98(4):043528.

Valcin, D., Bernal, J. L., Jimenez, R., Verde, L., and Wandelt, B. D. (2020). Inferring the age of the universe with globular clusters. *J. Cosmology Astropart. Phys.*, 2020(12):002.

van de Weygaert, R., Vegter, G., Edelsbrunner, H., et al. (2013). Alpha, Betti and the Megaparsec Universe: on the Topology of the Cosmic Web. *arXiv e-prints*, arXiv:1306.3640.

van den Busch, J. L., Wright, A. H., Hildebrandt, H., et al. (2022). KiDS-1000: cosmic shear with enhanced redshift calibration. *arXiv e-prints*, arXiv:2204.02396.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press.

van Waerbeke, L., Bernardeau, F., and Mellier, Y. (1999). Efficiency of weak lensing surveys to probe cosmological models. *A&A*, 342:15–33.

Van Waerbeke, L., Mellier, Y., Erben, T., et al. (2000). Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures. *A&A*, 358:30–44.

Wagoner, R. V., Fowler, W. A., and Hoyle, F. (1967). On the Synthesis of Elements at Very High Temperatures. *ApJ*, 148:3.

Walker, A. G. (1937). On Milne's Theory of World-Structure. *Proceedings of the London Mathematical Society*, 42:90–127.

Wasserman, L. (2018). Topological Data Analysis. *Annual Review of Statistics and Its Application*, 5(1):501–532.

Watkinson, C. A., Majumdar, S., Pritchard, J. R., and Mondal, R. (2017). A fast estimator for the bispectrum and beyond - a practical method for measuring non-Gaussianity in 21-cm maps. *MNRAS*, 472(2):2436–2446.

Welch, B., Coe, D., Diego, J. M., et al. (2022). A highly magnified star at redshift 6.2. *Nature*, 603(7903):815–818.

Wilczynska, M. R., Webb, J. K., Bainbridge, M., et al. (2020). Four direct measurements of the fine-structure constant 13 billion years ago. *Science Advances*, 6(17):eaay9672.

Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. (2020). H0LiCOW - XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes. *MNRAS*, 498(1):1420–1439.

Wright, A. H., Hildebrandt, H., Kuijken, K., et al. (2019). KiDS+VIKING-450: A new combined optical and near-infrared dataset for cosmology and astrophysics. *A&A*, 632:A34.

Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. (2020). KiDS+VIKING-450: Improved cosmological parameter constraints from redshift calibration with self-organising maps. *A&A*, 640:L14.

Xu, X., Cisewski-Kehe, J., Green, S. B., and Nagai, D. (2019). Finding cosmic voids and filament loops using topological data analysis. *Astronomy and Computing*, 27:34.

Yoo, J., Grimm, N., Mitsou, E., Amara, A., and Refregier, A. (2018). Gauge-invariant formalism of cosmological weak lensing. *J. Cosmology Astropart. Phys.*, 2018(4):029.

Zuntz, J., Paterno, M., Jennings, E., et al. (2015). CosmoSIS: Modular cosmological parameter estimation. *Astronomy and Computing*, 12:45–59.

Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., and Refregier, A. (2021). Cosmological forecast for non-Gaussian statistics in large-scale weak lensing surveys. *J. Cosmology Astropart. Phys.*, 2021(1):028.

Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127.

APPENDIX A

Testing the TREECORR code

The integration routine to calculate the $\Gamma^{(i)}$ described in Sect. 5.1.2.1 suffered from a bug for a long time, that unfortunately also persisted in the validation tests described in Sect. 5.1.2.3. We assumed that the integration worked because it passed all validation tests, so we conducted extensive tests of the TREECORR algorithm to ensure that the measurement of the 3pcf in N-body simulations works. All conducted tests verified that TREECORR measures the 3pcf with a good accuracy. This means that the tests are not very relevant to the results of this thesis. However, we still want to document the performed tests here to present an independent validation of the TREECORR algorithm.

We test the TREECORR algorithm by creating lognormal random fields on a 4096×4096 pixel grid describing a $10 \times 10\,\mathrm{deg}^2$ field. We first generate a GRF using a convergence power spectrum that corresponds to the *Euclid*-like SLICS. We then transform the GRF to lognormal fields, as described in Sect. 2.1.2.4, with the degree of lognormality $\alpha = 0.7$. We set the variance σ to be the same as the variance of the corresponding GRF.

We further construct a simple brute-force algorithm to measure both the shear and convergence 3pcf in the lognormal fields. Given a triangle configuration, the algorithm generates this triangle on the pixel grid. It then measures the 3pcf by moving the triangle around the field via translation and rotation and calculates the 3pcf as the average of all the resulting pixel triplets. In contrast to the TREECORR algorithm, this only measures a tiny sub-sample of the field's whole 3pcf. However, since lognormal fields are easy and cheap to generate, we can offset this by creating a larger number of fields.

We then calculate the convergence 3pcf ζ from the analytic equation (2.59) and compare it to both measurements from TREECORR, where we randomly select every 100-th pixel of the field, and the brute-force algorithm. In Fig. A.1, we can see a good agreement overall; on small scales TREECORR shows some strange behaviour, which is likely caused by the pixelisation of the shear grid. We further note that there are empty TREECORR-bins; this is due to the bin_slop parameter, which allows balls of the tree to go over the edges of a bin and causes some bins not to be filled when a pixelised grid is the input.

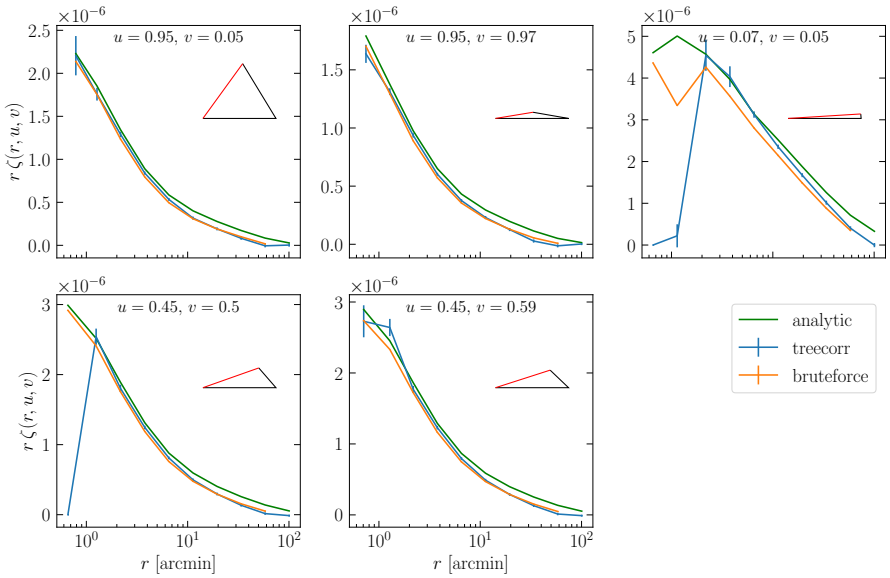


Figure A.1. – The convergence 3pcf ζ of a lognormal field from an analytic model (green), the TREECORR algorithm (blue) and a brute-force code (orange). In each panel, the triangle shape is fixed to the shape shown in the panel; the x axis describes the length of the red triangle side. For TREECORR, the error bars are errors on the mean of the 140 created lognormal realisations; in the brute-force code, we take the mean over 1000 lognormal fields.

We furthermore compare TREECORR to a bruteforce code on shear fields in Fig. A.2 and see an overall agreement that is similar to the one on convergence maps. There are some significant disagreements for degenerate triangles, which we attribute to pixelisation effects.

Lastly, we compare the results of the TREECORR algorithm when called on one line of sight of the MS with bin_slop parameters between 0.25 and 1. We find no discernible differences between the results; Secco et al. (2022a) noted that the expectation value of the 3pcf is relatively insensitive to the bin_slop parameter; however, the covariance is more strongly affected. As TREECORR does not play a role in our covariance calculations, this does not matter to us.

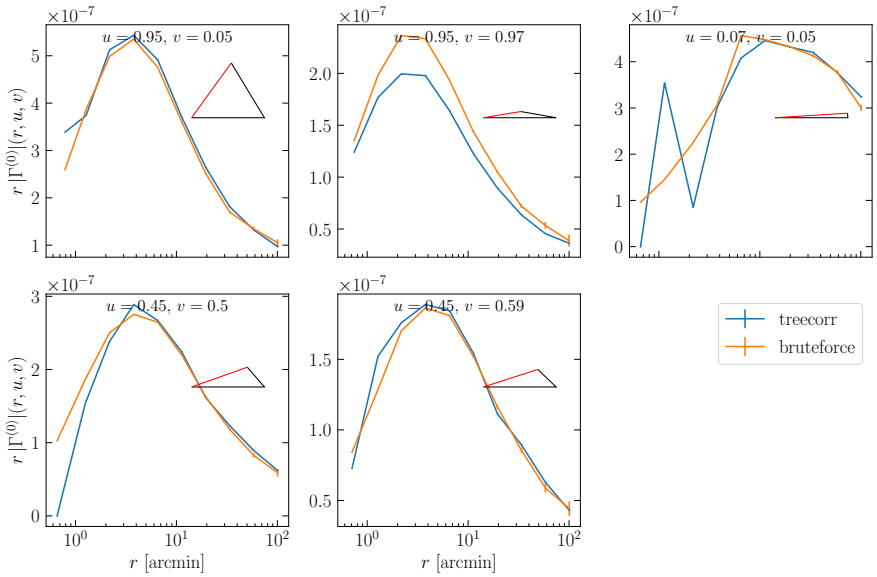


Figure A.2. – The absolute value of the first natural component $|\Gamma^{(0)}|$ of the shear 3pcf for a lognormal field from the TREECORR algorithm and a brute-force code. In each panel, the triangle shape is fixed to the shape shown in the panel; the x axis describes the length of the red triangle side. For TREECORR, the error bars are errors on the mean of the 140 created lognormal realisations; in the brute-force code, we take the mean over 1000 lognormal fields.

APPENDIX B

Measuring the third-order shear statistics in KiDS-1000

Although we are far from conducting a cosmological parameter analysis in a real survey, we want to show that we can measure the shear 3pcf in the KiDS-1000 survey and detect a third-order signal. To do this, we take the data release 4.1 of KiDS-1000 (Hildebrandt et al., 2021; Giblin et al., 2021) and remove the galaxies that have been flagged by the calibration method based on self-organising maps (Wright et al., 2020). For the remaining 21 million galaxies, we measure the shear 3pcf in 16^3 bins from $\theta_{\rm min} = 0.'1$ to $\theta_{\rm max} = 150'$. We further calculate a model 3pcf at the best-fit cosmology for KiDS-1000 determined by van den Busch et al. (2022). The results can be seen in Fig. B.1. We can see that even with the errors disregarding any sample variance, each individual bin contains no discernible signal and is dominated by noise. This is not entirely surprising, as the measured third-order shear signal is divided up into 8×4096 bins, so the signal in each bin is extremely small. However, these measured correlation functions are a good basis for calculating the third-order aperture statistics.

We convert the measured 3pcf to third-order aperture masses with aperture radii $4', 8', 16'$ and $32'$. The results can be seen in Fig. B.2. We see that the model at the best-fit cosmology of KiDS-1000 (van den Busch et al., 2022) agrees with the measurement on all scales within the error bars. We also note that we measure no significant B-modes in the survey. This exciting result shows that we can be optimistic about our ability to analyse third-order shear statistics in Stage-III surveys soon. We detect the cosmic shear signal with about

$$\sqrt{\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm meas} C^{-1} \left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm meas}} \sigma \approx 6.7\sigma \ , \tag{B.1}$$

where $\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm meas}$ are the third-order aperture statistics we measure in KiDS-1000 and C is the covariance matrix determined from the analytical model (Sect. 5.2.2) for the redshift distribution and survey area of KiDS-1000. For the deviation from the best-fit model, we

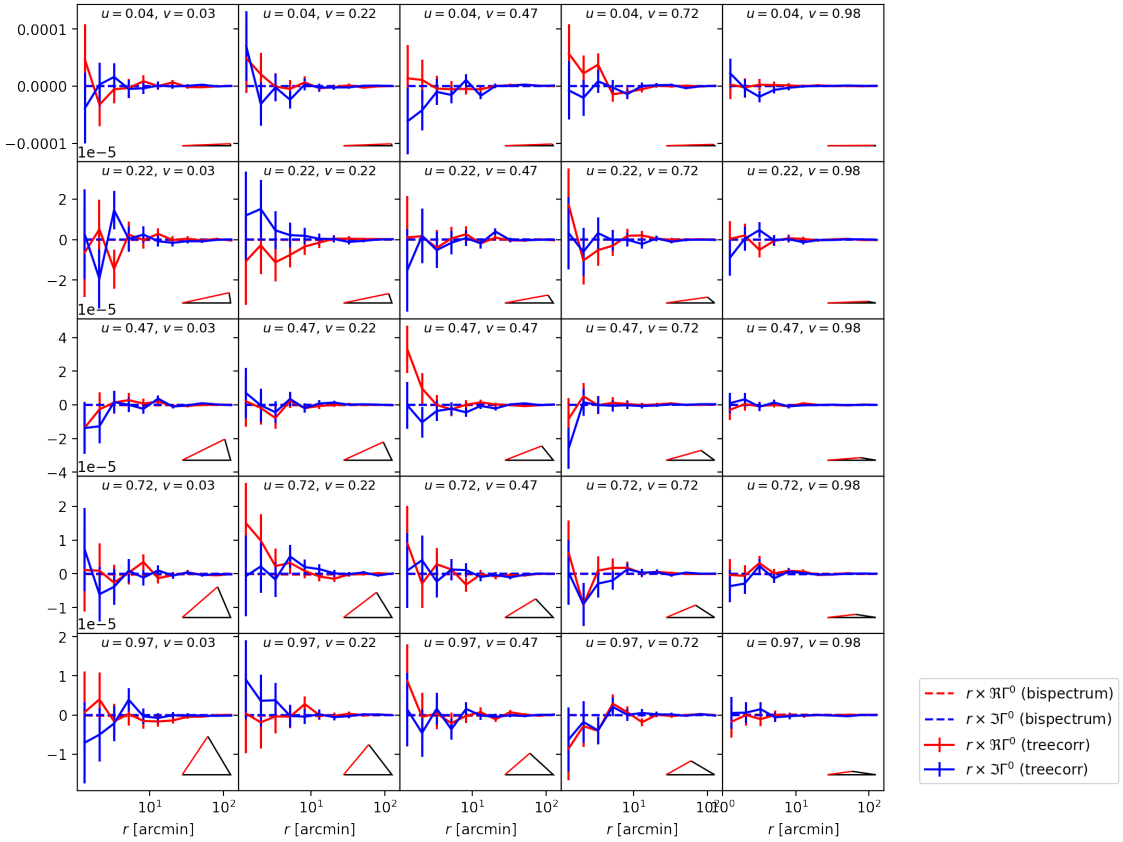


Figure B.1. – The first natural component of the shear three-point correlation functions as measured in KiDS-1000 using TreeCorr. For comparison, we show a model at the best-fit cosmology of KiDS-1000 determined by van den Busch et al. (2022). Each panel corresponds to one fixed triangle shape; the u and v values are listed at the top, and the corresponding shape of the triangle is shown on the bottom right. The x-axis corresponds to the length of the red side of the triangle. We show both the real part (red) and the imaginary part (blue) for the model (dashed line) and the simulations (solid line). The error bars are estimated from the shape noise of all galaxies in a bin and disregard sample variance.

measure

$$\chi^2/\text{d.o.f.} = \frac{\left(\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm meas} - \left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm model}\right) C^{-1} \left(\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm meas} - \left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm model}\right)}{20} \approx 0.29 \,, \tag{B.2}$$

where the degrees of freedom of our data vector are d.o.f. = 20 and $\left\langle \mathcal{M}_{\rm ap}^3 \right\rangle_{\rm model}$ are the third-order aperture statistics modelled at the best-fit cosmology of van den Busch et al. (2022). However, these numbers should not be taken at face value, as our covariance model is imperfect, and we can not rule out that systematic effects partly cause the measured signal.

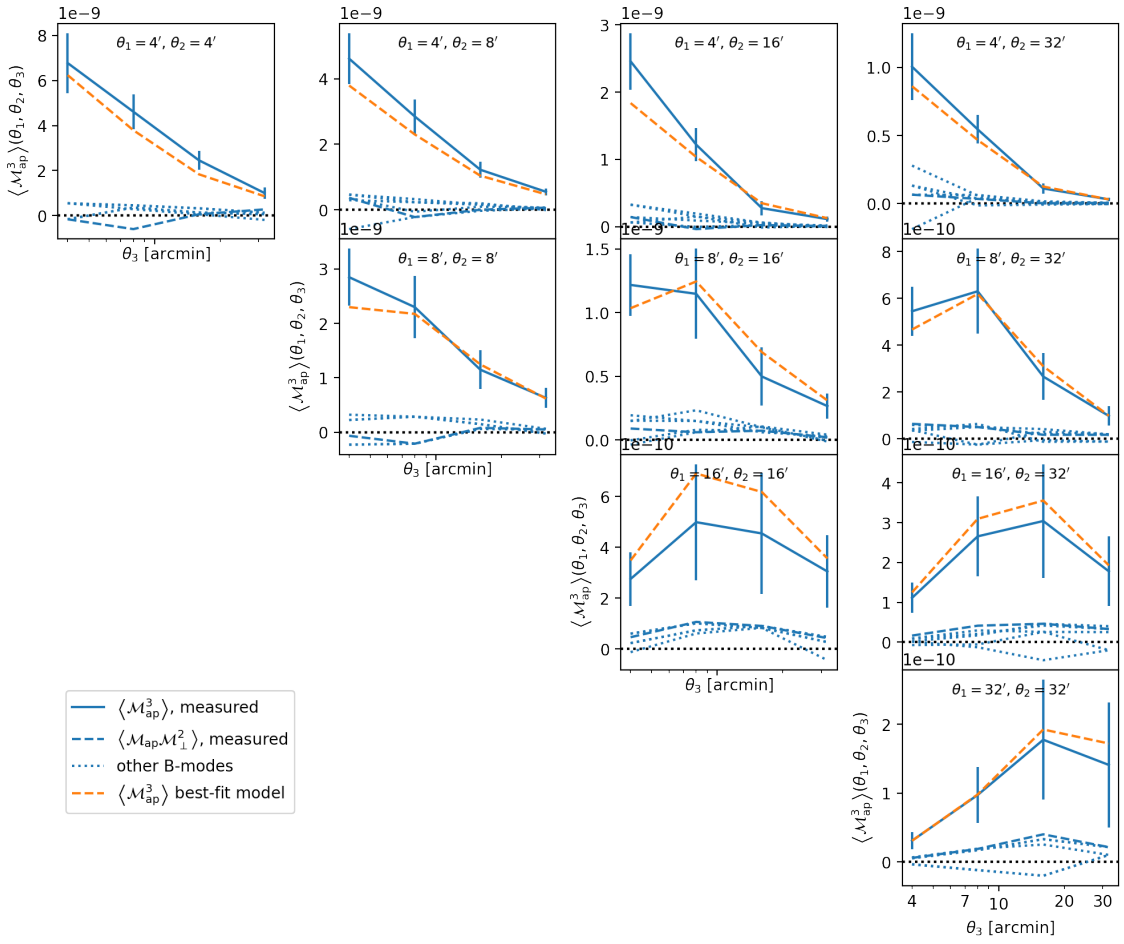


Figure B.2. – The third-order aperture mass statistics as measured in KiDS-1000 using TREECORR. For comparison, we show our model at the best-fit cosmology of KiDS-1000. The error bars are calculated using the theoretical model described in Sect. 5.2.

APPENDIX C

Published papers

C.1. Persistent homology in cosmic shear: Constraining parameters with topological data analysis

A&A 648, A74 (2021)
https://doi.org/10.1051/0004-6361/202039048

Astronomy & Astrophysics

Persistent homology in cosmic shear: Constraining parameters with topological data analysis

Sven Heydenreich[1], Benjamin Brück[2], and Joachim Harnois-Déraps[3,4]

[1] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
e-mail: sven@astro.uni-bonn.de
[2] University of Copenhagen, Department of Mathematical Sciences, Universitetsparken 5, 2100 Copenhagen, Denmark
[3] Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK
[4] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

Received 28 July 2020 / Accepted 8 December 2020

ABSTRACT

In recent years, cosmic shear has emerged as a powerful tool for studying the statistical distribution of matter in our Universe. Apart from the standard two-point correlation functions, several alternative methods such as peak count statistics offer competitive results. Here we show that persistent homology, a tool from topological data analysis, can extract more cosmological information than previous methods from the same data set. For this, we use persistent Betti numbers to efficiently summarise the full topological structure of weak lensing aperture mass maps. This method can be seen as an extension of the peak count statistics, in which we additionally capture information about the environment surrounding the maxima. We first demonstrate the performance in a mock analysis of the KiDS+VIKING-450 data: We extract the Betti functions from a suite of N-body simulations and use these to train a Gaussian process emulator that provides rapid model predictions; we next run a Markov chain Monte Carlo analysis on independent mock data to infer the cosmological parameters and their uncertainties. When comparing our results, we recover the input cosmology and achieve a constraining power on $S_8 \equiv \sigma_8 \sqrt{\Omega_\mathrm{m}/0.3}$ that is 3% tighter than that on peak count statistics. Performing the same analysis on 100 deg^2 of *Euclid*-like simulations, we are able to improve the constraints on S_8 and Ω_m by 19% and 12%, respectively, while breaking some of the degeneracy between S_8 and the dark energy equation of state. To our knowledge, the methods presented here are the most powerful topological tools for constraining cosmological parameters with lensing data.

Key words. gravitational lensing: weak – cosmological parameters – methods: data analysis

1. Introduction

The Λ cold dark matter (ΛCDM) model is mostly regarded as the standard model of cosmology and has been incredibly successful in explaining and predicting a large variety of cosmological observations using only six free parameters. As of now, the tightest constraints on these cosmological parameters have been placed by observations of the cosmic microwave background (CMB, Planck Collaboration VI 2020). However, the next generation of surveys, such as the *Rubin* Observatory[1] (Ivezic et al. 2008), *Euclid*[2] (Laureijs et al. 2011) and the Nancy Grace Roman Space Telescope[3] (RST, Spergel et al. 2013), promises to improve on these constraints and reduce parameter uncertainties to the sub-percent level. This is particularly interesting in light of tensions arising between observations of the early Universe (CMB) and the late Universe at a redshift of $z \lesssim 2$. These tensions are most notable in the Hubble constant H_0 (Riess et al. 2019) and the parameter $S_8 = \sigma_8 \sqrt{\Omega_\mathrm{m}/0.3}$ (Joudaki et al. 2020; Abbott et al. 2020), where Ω_m is the matter density parameter and σ_8 characterises the normalisation of the matter power spectrum. Future analyses will show whether these tensions are a statistical coincidence, the manifestation of unknown systematics or evidence for new physics.

In the last decade, weak gravitational lensing has proven to be an excellent tool for studying the distribution of matter and constraining the cosmological parameters that describe our Universe. In particular, tomographic shear two-point correlation functions (Kaiser 1992) and derived quantities such as Complete Orthogonal Sets of E- and B-mode Integrals (COSEBIs, Schneider et al. 2010) and band powers (van Uitert et al. 2018) have been applied with great success to cosmic shear data (see Kilbinger et al. 2013; Heymans et al. 2013; Joudaki et al. 2017 for an analysis of CFHTLenS, Hildebrandt et al. 2020; Wright et al. 2020; Asgari et al. 2020 for the Kilo Degree Survey, Troxel et al. 2018 for the Dark Energy Survey and Hikage et al. 2019; Hamana et al. 2020 for the Hyper Suprime Camera Survey). However, while two-point statistics provide an excellent tool for capturing the information content of Gaussian random fields, the gravitational evolution of the matter distribution becomes non-linear in high-density regions, and additional methods are needed to extract the information residing in the non-Gaussian features that are formed therein. The demand for these new methods rises with the ever-increasing amount and quality of available data.

In addition to three-point correlation functions (Schneider & Lombardi 2003; Fu et al. 2014), which form a natural extension to two-point correlation functions, a variety of alternative statistics promise to improve cosmological parameter constraints (Zürcher et al. 2021), including peak statistics (Liu et al. 2015a,b; Kacprzak et al. 2016;

[1] https://www.lsst.org/
[2] https://www.euclid-ec.org/
[3] https://roman.gsfc.nasa.gov/

Martinet et al. 2018, hereafter M+18), density split statistics (Gruen et al. 2018), convergence probability distribution functions (Liu & Madhavacheril 2019), shear clipping (Giblin et al. 2018), the scattering transform (Cheng et al. 2020), and Minkowski functionals (Petri et al. 2015; Marques et al. 2019; Parroni et al. 2020).

In this article, we demonstrate how 'persistent homology' can be used to analyse the data provided by weak gravitational lensing surveys. This non-linear statistic combines the information residing in Minkowski functionals and in peak statistics in a natural way, and supplements these by further capturing information about the environment surrounding the topological features. In the last two decades, persistent homology has been successfully applied in a large variety of fields involving topological data analysis: Among others, it has been used to analyse the spread of contagious diseases (Lo & Park 2018), to assess self-similarity in geometry (MacPherson & Schweinhart 2012) and to identify a new subgroup of breast cancer (Nicolau et al. 2011); for more examples, we refer to Otter et al. (2017) and references therein.

In cosmology, persistent homology has already been used to study the topology of the cosmic web (Sousbie 2011; van de Weygaert et al. 2013), of the interstellar magnetic fields (Makarenko et al. 2018) and the reionisation bubble network (Elbers & van de Weygaert 2019). Furthermore, it was shown to be an effective cosmic void finder (Xu et al. 2019). Additionally, a specific summary of the information contained in persistent homology, the 'Betti numbers', has emerged as a powerful tool to analyse both the cosmic web (Pranav et al. 2017) and Gaussian random fields (Pranav et al. 2019a, hereafter P+19). In the latter case, Betti numbers have been shown to provide a higher information content than both Euler characteristics and Minkowski functionals (Pranav et al. 2019b). In particular, they are very effective at detecting and quantifying non-Gaussian features in fields such as the CMB temperature map (compare P+19).

While several of the above-mentioned papers have shown that Betti numbers are sensitive to cosmological parameters, their results were only qualitative so far. To our knowledge, the current article is the first to quantify the power of Betti numbers for constraining cosmological parameters. Furthermore, our work differs from the previous ones in that we use 'persistent' Betti numbers, which further take into account the environment surrounding the topological features (for the definitions and further explanations, see Sect. 2.3). We show that this leads to a significant improvement in constraining power compared to the 'non-persistent' Betti numbers that have so far been used in cosmology.

In our analysis, we constrain cosmological parameters with a Markov chain Monte Carlo (MCMC) sampler, and therefore need to efficiently compute Betti functions over a broad range of values. Since we are not aware of a way to model Betti functions for non-Gaussian fields, we instead rely on Gaussian process regression (see e.g., Gelman et al. 2004), a machine learning tool that probabilistically predicts a given statistic when only small sets of training data are available. This procedure is not restricted to Betti functions and has been used in cosmology for a number of other statistical methods (Heitmann et al. 2014; Liu & Madhavacheril 2019; Burger et al. 2020; Mootoovaloo et al. 2020). Since persistent homology is particularly efficient in summarising and compressing the topological structure of large data sets, it is well suited for interacting with machine learning algorithms (Bresten & Jung 2019).

In this work, we train a Gaussian process regressor on a suite of KiDS+VIKING-450-like (Wright et al. 2019, hereafter KV450-like) N-body simulations to predict Betti functions for arbitrary cosmological parameters within a wide training range. All simulations are part of the w cold dark matter (wCDM) model of cosmology, meaning that they extend the standard model by allowing the equation of state of dark energy to vary. Calibrating our covariance matrix from a distinct ensemble of fully independent simulations, we then perform an MCMC analysis on mock data and recover the fiducial cosmological parameters of the simulation. We further show that persistent Betti functions are able to constrain cosmological parameters better than peak statistics, whose performance is similar to the one of tomographic cosmic shear (Kacprzak et al. 2016; M+18) and slightly better than that of Minkowski functionals (Zürcher et al. 2021).

We finally carry out a mock analysis based on *Euclid*-like simulations and find an even larger increase in constraining power in this setup. Thus, persistent Betti numbers promise to substantially improve the constraining power of weak gravitational lensing surveys, especially when they are used in combination with other probes.

This paper is organised as follows: In Sect. 2 we give a brief overview of the N-body simulations used in the signal calibration and in the estimation of the covariance matrix (Sect. 2.1), of the aperture mass maps reconstruction (Sect. 2.2), of the theory underlying persistent homology statistics (Sect. 2.3) and Gaussian process regression emulation (Sect. 2.4). An explanation of the peak count statistics can be found in Sect. 2.5. We present the results of our KV450-like and *Euclid*-like analyses in Sects. 3 and 4, respectively, and discuss them in Sect. 5.

2. Methods and numerical data products

Throughout this work, we assume the standard weak gravitational lensing formalism, a review of which can be found in Bartelmann & Schneider (2001).

2.1. Weak lensing simulations

So far, we cannot analytically compute the Betti functions that describe cosmic shear data due to their highly non-linear nature. Instead, we rely on numerical simulations[4], namely the Scinet LIghtCone Simulations (SLICS, Harnois-Déraps et al. 2018; Harnois-Déraps & van Waerbeke 2015) and the cosmo-SLICS (Harnois-Déraps et al. 2019, hereafter H+19) for their evaluation. All simulations were performed in a flat wCDM framework, which we will assume throughout the scope of this paper.

First, we extracted the cosmology dependence with the cosmo-SLICS, a suite of cosmological N-body simulations in which the matter density Ω_m, the parameter S_8, the Hubble constant h and the parameter for the dark energy equation of state w_0 are sampled at 26 points in a Latin hyper-cube (see Table A.1 for the exact list). At each cosmology, a pair of N-body simulations were evolved to redshift $z = 0$ with 1536^3 particles in a box of $505\,h^{-1}$ Mpc, and subsequently ray-traced multiple times to yield 50 pseudo-independent light-cones that cover each $10 \times 10\,\mathrm{deg}^2$. The initial conditions were chosen in order to suppress most of the sample variance when averaging a statistic over the pair. (For more details on the cosmo-SLICS, we refer the reader to H+19).

Second, we used the SLICS to estimate the covariance of the Betti functions. These consist of a set of 126 fully independent

[4] All simulations products used in this paper can be made available upon request; see http://slics.roe.ac.uk

N-body simulations[5] conducted with $\Omega_\mathrm{m} = 0.2905$, $\Omega_\mathrm{b} = 0.0473$, $h = 0.6898$, $\sigma_8 = 0.826$ and $n_\mathrm{s} = 0.969$. These were also ray-traced into $10 \times 10\,\mathrm{deg}^2$ light-cones, in a format otherwise identical to the cosmo-SLICS.

We took the KV450 data set (Wright et al. 2019) as an example of a current Stage-III weak lensing survey and created mock data sets with similar properties. Due to the box size of our simulations, the full KV450 survey footprint cannot be fitted onto a single light-cone. Instead we split the survey into 17 tiles following the setup presented in Appendix A.3 of Harnois-Déraps et al. (2018) and computed the (simulated) shear signal at the exact positions of the KV450 galaxies, repeating the process for ten light-cones (out of the 50 available) for each cosmo-SLICS pair and for the 126 SLICS realisations. We further used the observed ellipticities to simulate the shape noise, and the galaxy redshifts were randomly selected such that the cumulative redshift distribution follows the fiducial 'direct calibration' method (DIR) described in Hildebrandt et al. (2020). This way, any effect that the galaxy density, shape noise, and survey footprint may have on the Betti functions is infused in the simulations as well.

In Sect. 4 we further used a separate set of simulations in which the redshift distribution and galaxy number density has been modified to match that of future weak lensing experiments. These *Euclid*-like catalogues are also based on the SLICS and cosmo-SLICS, and we provide more details about them in that section. We note that in neither of the simulated surveys do we split the galaxy catalogues with redshift selection; we leave tomographic analysis for future work.

2.2. Maps of aperture mass

We performed our topological analysis on data obtained from weak gravitational lensing surveys, more precisely on maps of aperture mass (Schneider 1996), which reconstruct the lensing convergence κ inside an aperture filter and are hence directly related to the projected mass density contrast. There are several alternative ways to reconstruct convergence maps (see, e.g., Kaiser & Squires 1993; Seitz & Schneider 2001; Jeffrey et al. 2018, and references therein), however, these suffer from the so-called mass-sheet degeneracy: Even under ideal circumstances, κ can only be determined up to a constant. The introduction of a uniform mass sheet κ_0 can change the extracted signal-to-noise ratio (S/N) in a way that does not reflect any physical meaning, hence we choose to perform our analysis on aperture mass maps, which are invariant under this effect.

2.2.1. Theoretical background

The aperture mass is obtained from a weak lensing catalogue as

$$M_\mathrm{ap}(\boldsymbol{\theta}) = \int \mathrm{d}^2\theta' Q(|\boldsymbol{\theta}'|)\gamma_\mathrm{t}(\boldsymbol{\theta}';\boldsymbol{\theta}), \tag{1}$$

where the filter function Q is computed via

$$Q(\theta) = \frac{2}{\theta^2}\int_0^\theta \mathrm{d}\theta'\,\theta' U(\theta') - U(\theta) \tag{2}$$

and U is a compensated filter with $\int \mathrm{d}\theta\,\theta U(\theta) = 0$ (Schneider 1996). The tangential shear at position $\boldsymbol{\theta}'$ with respect to $\boldsymbol{\theta}$, $\gamma_\mathrm{t}(\boldsymbol{\theta}';\boldsymbol{\theta})$, is calculated as

$$\gamma_\mathrm{t}(\boldsymbol{\theta}';\boldsymbol{\theta}) = -\Re\left[\gamma(\boldsymbol{\theta}')\frac{(\boldsymbol{\theta}'^* - \boldsymbol{\theta}^*)^2}{|\boldsymbol{\theta}' - \boldsymbol{\theta}|^2}\right], \tag{3}$$

[5] The full SLICS ensemble contains 820 realisations, however we only use a sub-sample of 126 in this work.

where $\gamma(\boldsymbol{\theta}')$ is the (complex) shear at position $\boldsymbol{\theta}'$. We note that both the shear γ and the angular position $\boldsymbol{\theta}$ are interpreted as complex quantities in the above expression. In reality, we do not measure a shear field directly, but rather ellipticities ϵ of an ensemble of galaxies n_gal, which are related to the shear and a measurement noise term ϵ_n as

$$\epsilon = \frac{\epsilon_\mathrm{n} + g}{1 + \epsilon_\mathrm{n} g^*} \sim \epsilon_\mathrm{n} + \gamma, \tag{4}$$

where $g = \gamma/(1-\kappa)$ is the reduced shear and the last approximation holds in the weak shear limit (see e.g., Bartelmann & Schneider 2001). This transforms the integral of Eq. (1) into a sum

$$M_\mathrm{ap}(\boldsymbol{\theta}) = \frac{1}{n_\mathrm{gal}}\sum_i Q(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|)\epsilon_\mathrm{t}(\boldsymbol{\theta}_i;\boldsymbol{\theta}), \tag{5}$$

where $\epsilon_\mathrm{t}(\boldsymbol{\theta}_i;\boldsymbol{\theta})$ is the tangential ellipticity defined in analogy to Eq. (3).

Following M+18, we computed the noise in the aperture mass as

$$\sigma\left(M_\mathrm{ap}(\boldsymbol{\theta})\right) = \frac{1}{\sqrt{2}n_\mathrm{gal}}\sqrt{\sum_i |\epsilon(\boldsymbol{\theta}_i)|^2 Q^2(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|)} \tag{6}$$

and calculated the S/N at a position $\boldsymbol{\theta}$ as

$$\frac{S}{N}(\boldsymbol{\theta}) = \frac{\sqrt{2}\sum_i Q(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|)\,\epsilon_\mathrm{t}(\boldsymbol{\theta}_i;\boldsymbol{\theta})}{\sqrt{\sum_i |\epsilon(\theta_i)|^2 Q^2(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|)}}. \tag{7}$$

Both numerator and denominator of Eq. (7) can be expressed as a convolution and can therefore be computed via a Fast Fourier-Transform (FFT), significantly decreasing the required computation time (Unruh et al. 2020). For our analysis, we used the following filter function (see Schirmer et al. 2007; M+18):

$$Q(\theta) = \left[1 + \exp\left(6 - 150\frac{\theta}{\theta_\mathrm{ap}}\right) + \exp\left(-47 + 50\frac{\theta}{\theta_\mathrm{ap}}\right)\right]^{-1} \times \left(\frac{\theta}{x_\mathrm{c}\theta_\mathrm{ap}}\right)^{-1}\tanh\left(\frac{\theta}{x_\mathrm{c}\theta_\mathrm{ap}}\right). \tag{8}$$

This function was designed to efficiently follow the mass profiles of dark matter haloes, which we model according to Navarro et al. (1997, hereafter NFW); since most of the matter is located within dark matter haloes, the function is well suited to detect peaks in the matter distribution. Here, θ_ap is the aperture radius and x_c represents the concentration index of the NFW profile. Following M+18, we set $x_\mathrm{c} = 0.15$, which is a good value for detection of galaxy clusters (Hetterscheidt et al. 2005).

2.2.2. Numerical implementation

The S/N defined in Eq. (7) is highly sensitive to the noise properties of the galaxy survey, and it is therefore critical to reproduce exactly the galaxy number density and the intrinsic shape noise of the data, as well as the overall survey footprint (see e.g., M+18). This motivates our use of KV450 mosaic simulations, which reproduce all of these quantities exactly. For each simulated galaxy in the catalogue, we randomly rotated the observed ellipticity of the corresponding real galaxy and added it to the simulated shear following the linear approximation in Eq. (4).

To calculate the aperture mass maps in a computationally inexpensive way, we distributed the galaxies on a grid of pixel size 0.′6. For each pixel, we computed the sum of all respective galaxy ellipticities $\sum_i \epsilon_i$ as well as the sum of their squared absolute ellipticities $\sum_i |\epsilon_i|^2$. The distribution of galaxies on a grid slightly shifts their positions, which introduces a small error in the computed quantity ϵ_t when the value $|\boldsymbol{\theta}_i - \boldsymbol{\theta}|$ of Eq. (7) is comparable to the size of a pixel, but it enables a significantly faster computation[6]. Moreover, since the convolutions arising in both the numerator and denominator of Eq. (7) are linear in ϵ and $|\epsilon|^2$, respectively, computing Eq. (7) for individual galaxies or for whole pixels does not make a difference, and the latter is more efficient. Finally, we chose a filter radius of $\theta_{\rm ap} = 12.'5$ in this work, however the choice of this parameter could be revisited in order to optimise the cosmological constraints.

In addition to exhibiting an irregular shape, our KV450-like footprint is affected by internally masked regions, for example by the removal of bright foreground stars, saturated galaxies or satellite tracks. Each pixel that contains zero galaxies, which is true in particular for masked regions, was treated as masked. For the calculation of the aperture mass maps, these masked shear pixels were treated as having a shear of 0, and we subsequently masked pixels in the signal-to-noise maps for which at least 50% of the aperture was made of masked shear pixels; masked S/N pixels were assigned the value $-\infty$. As our simulated data exactly traces the KV450 footprint, our cosmological parameter analysis is not affected by the masked regions.

2.3. Topological data analysis

This section describes the methods employed to conduct the topological analysis of the aperture mass maps detailed in Sect. 2.2.2, from the mathematical background to a full description of our numerical implementation.

Our analysis is based on the study of topological features in the S/N maps extracted from mock weak lensing data. More particularly, we are interested in studying how these relate to their environment in order to access non-Gaussian information contained in the correlation between different scales of the shear field. The idea of our approach is as follows: First, we take the survey area X and we remove from it all pixels that have S/N value above a threshold t. The result is a (topological) space that can be seen as a part of X (see Fig. 1). Next, we count the number of connected components[7] and holes in this space. These are called the 'Betti numbers' β_0 and β_1. Lastly, we analyse how these numbers change as we vary the threshold t.

Figure 2 presents an example of a sequence of spaces obtained in this way. It shows the filtered aperture mass maps for nine values of t, obtained from a square-shaped zoom-in region of the full survey area X. Inspecting the panels from top-left to bottom-right, we see the gradual recovery of the full aperture mass map as the threshold t increases, starting from the lower values. One sees that a local minimum of the S/N map corresponds to a connected component appearing at some point and vanishing later on as it gets absorbed by an 'older' feature; a maximum corresponds to a hole that first appears and later gets filled in. Analysing how these features are created and vanish again allows us not only to count the corresponding extrema but to take into account their environment as well, thus obtaining information about the large-scale structure.

2.3.1. Homology of excursion sets and local extrema

Mathematically, the ideas described above can be expressed as follows: Let $f: X \to \mathbb{R}$ be a map from a topological space X to the reals. In our context, X will be the observed survey area, which we interpret as a subspace of the celestial sphere S^2, and f will be the S/N of the aperture mass, $f(\boldsymbol{\theta}) = S/N(\boldsymbol{\theta})$, defined in Eq. (7). Taking sublevel sets of the map f (i.e. the portion of f with values less than some maximal value t) yields a sequence of subspaces of X: For $t \in \mathbb{R}$, define the 'excursion set' $X_t = \{x \in X \mid f(x) \leq t\}$.

To simplify notation, assume that f is bounded. If we choose $t_1 \leq \ldots \leq t_k > \sup(f)$, we obtain a 'filtration' of the space X, which is defined as a sequence of subspaces with:

$$X_1 \subseteq X_2 \subseteq \cdots \subseteq X_k = X, \tag{9}$$

where we write $X_i = X_{t_i}$ in the above expression to make the notation more compact. Our aim is to extract cosmological information[8] from the distribution of aperture mass (represented by the map f) by studying topological invariants of the sequence $\mathbb{X} = (X_i)_i$.

A 'topological invariant' of a space is a mathematical object (e.g., a number) that does not change if we continuously perturb it, for example by stretching or bending. The case that is relevant for us are the Betti numbers. As mentioned above, these count the number of connected components and holes of a space. If we compute the Betti numbers of our excursion sets X_i, we obtain information about the number of local extrema of the map f: The minima correspond to connected components appearing at some point of the filtration $\mathbb{X}$ and vanishing later on; the maxima correspond to holes (see Fig. 2). The time it takes for the connected components or holes to vanish again is related to the relative height of the corresponding extremum, and therefore contains information about the environment.

From a more formal point of view, the nth Betti number $\beta_n(Y)$ of a space Y is computed from its nth homology group $H_n(Y)$. Both homology groups and Betti numbers are important and well-studied invariants from algebraic topology. For an introduction to algebraic topology that is geared towards its applications in the sciences, see Ghrist (2014); a more rigorous introduction to this area of mathematics can also be found in Hatcher (2002). In our context, the homology group $H_n(Y)$ is a vector space that we can associate with Y and $\beta_n(Y)$ is its dimension. The important point for us is that Betti numbers and homology groups translate the geometric (or rather topological) problem of counting connected components and holes into a question about linear algebra that can be efficiently solved by a computer (see Sect. 2.3.4).

We note that other topological invariants have been successfully used in cosmology, including the Euler characteristic[9] $\chi(Y)$. Following Adler (1981), the Euler characteristic can be used to study real-valued random fields such as the CMB (Hamilton et al. 1986; Parroni et al. 2020). It is given by the alternating sum of the Betti numbers; in particular, if Y is a

[6] This approximation will likely need to be revisited in future analyses with increased accuracy requirement.

[7] A connected component is a cluster of pixels connected by an edge or a corner.

[8] We note that the theory described in this section works exactly the same e.g., for analysing a grey-shade picture where X would be a rectangle and f the intensity at each point; a particular strength of the theory is that it also easily generalises to higher-dimensional data.

[9] If Y is a d-dimensional manifold, then $\chi(Y)$ is, up to a scalar factor, given by the top-dimensional Minkowski functional.

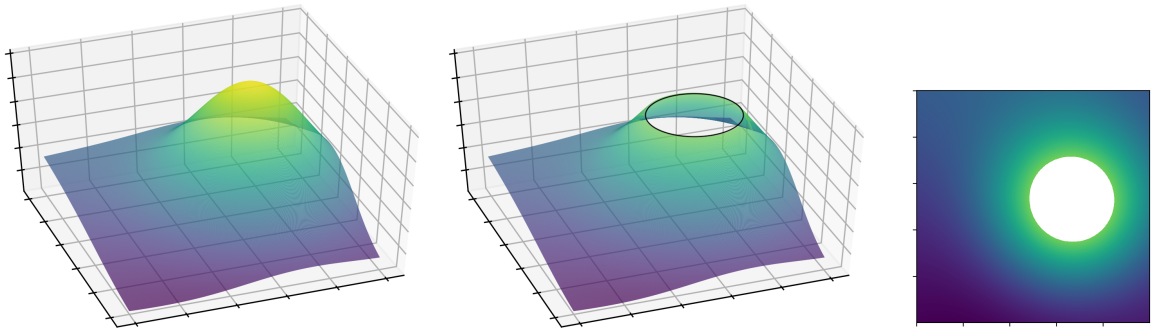
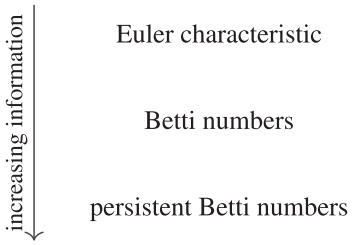


Fig. 1. Visualisation of an excursion set. The original map (in this case a single peak) is depicted on the left. The middle plot depicts the same peak after a threshold is applied to the map, which cuts off the summit. The result seen from above is depicted on the right, which in this case has one hole (in white) and one connected component (in colour). Varying the threshold generates the full filtration.

subset of $\mathbb{R}^2$, we have $\chi(Y) = \beta_0(Y) - \beta_1(Y)$. Hence, one can compute the Euler characteristic from the Betti numbers, but the latter contain strictly more information (see also the discussion in Pranav et al. 2019b), motivating our choice for this current data analysis. Additionally, 'persistent' Betti numbers, which we describe in the next subsection, provide us with even more information. The information content extracted from different topological methods can be summarised as follows:



This is quantified in Sect. 3.2 (see in particular Fig. 5).

2.3.2. Persistent homology

As discussed in the last section, one approach to studying the topological properties of the sequence of excursion sets $\mathbb{X} = (X_i)_i$ is to analyse the sequence of Betti numbers $\beta_n(X_i)$, for $n = 0, 1$. Algebraically, this amounts to studying the sequence of vector spaces $H_n(X_i)$ and computing each of their dimensions individually. However, one can do better: The inclusion maps $X_i \to X_{i+1}$ presented in Eq. (9) induce maps $H_n(X_i) \to H_n(X_{i+1})$, and these provide us with additional information. The sequence of all $H_n(X_i)$ together with the connecting maps form what is called a 'persistence module' that we will write as $H_n(\mathbb{X})$, and which describes the persistent homology of $\mathbb{X}$. For an introduction to persistent homology that focuses on its applications, we refer to Otter et al. (2017); a broader overview and further background material can be found in Oudot (2015).

The data of the persistence module $H_n(\mathbb{X})$ can be summarised in its persistence diagram $\mathrm{Dgm}(H_n(\mathbb{X}))$. This is a collection of half-open intervals $[t_i, t_j)$, where each interval[10] can be interpreted as a feature[11] that is 'born at level' t_i and 'dies' at level t_j. For example, an element in $\mathrm{Dgm}(H_0(\mathbb{X}))$ is an interval of the form $[t_i, t_j)$ and it corresponds to a connected component that appears at level t_i and merges with another component at level t_j. Similarly, elements of $\mathrm{Dgm}(H_1(\mathbb{X}))$ correspond to holes that appear at t_i and get filled at time t_j. The filtration sequence presented in Fig. 2 exhibits such features, some of which we have highlighted in the panels: The magenta ellipses present a connected component corresponding to a minimum of the signal-to-noise map; the orange and red ellipses each highlight a hole corresponding to a maximum. As the sequence evolves, these features appear and disappear, giving rise to intervals.

To conduct our analysis, we study the nth persistent Betti number (or rank invariant) of $\mathbb{X}$. This is defined as the number of features that are born before t and die after t', and can be extracted from the persistence diagram $\mathrm{Dgm}(H_n(\mathbb{X}))$ by

$$\beta_n\mathbb{X}(t, t') = \#\left\{[t_i, t_j) \in \mathrm{Dgm}(H_n(\mathbb{X})) \mid t_i \le t \le t' < t_j\right\}, \tag{10}$$

where, $\#\{\ldots\}$ denotes the number of elements of the set $\{\ldots\}$. We can consider these as functions from a subset of $\mathbb{R}^2$ to $\mathbb{R}$; if we want to emphasise this point of view, we will also call them *Betti functions*. We note that when $t' = t$ we recover the non-persistent Betti numbers:

$$\beta_n\mathbb{X}(t_i, t_i) = \beta_n(X_i). \tag{11}$$

However, the persistent Betti numbers contain strictly more information than the sequence of regular Betti numbers. In particular, if one knows all values of $\beta_n\mathbb{X}$, one can recover the entire persistence diagram $\mathrm{Dgm}(H_n(\mathbb{X}))$, while this is not possible using the sequence $(\beta_n(X_i))_i$. Intuitively speaking, the persistent Betti functions do not only provide us with information about how many connected components or holes we have in each excursion set X_i, but they also tell us how long each such feature persists throughout the filtration. As explained above, the features correspond to local minima and maxima of the S/N map. Knowing about their lifetimes provides information about the relative height of these extrema and about their entanglement: Two completely separated peaks can be distinguished from two peaks sitting together on the summit of a region of high S/N.

2.3.3. Masks and relative homology

As explained in Sect. 2.1, our S/N maps contain masked regions where we do not have sufficient information about the aperture mass (and their pixel values are set to be constant $-\infty$). In order to incorporate these in our analysis, we work with relative homology. The idea of using relative homology to study masked data was also used in P+19, where the authors give further interpretations of these groups. Other occurrences of relative homology in the persistent setting can be found in Pokorny et al. (2016) and Blaser & Brun (2019). Cf. also extended persistence and variants of it (Sect. 4 of Edelsbrunner & Harer 2008; de Silva et al. 2011).

[10] Such a collection of intervals is called a 'diagram' because it can be visualised as an actual diagram in the plane: For every interval $[t_i, t_j)$, one can draw a point in $\mathbb{R}^2$ with coordinates (t_i, t_j). An example of this can be found in Fig. B.1.

[11] Formally, these features represent basis elements of the vector spaces in the persistence module.

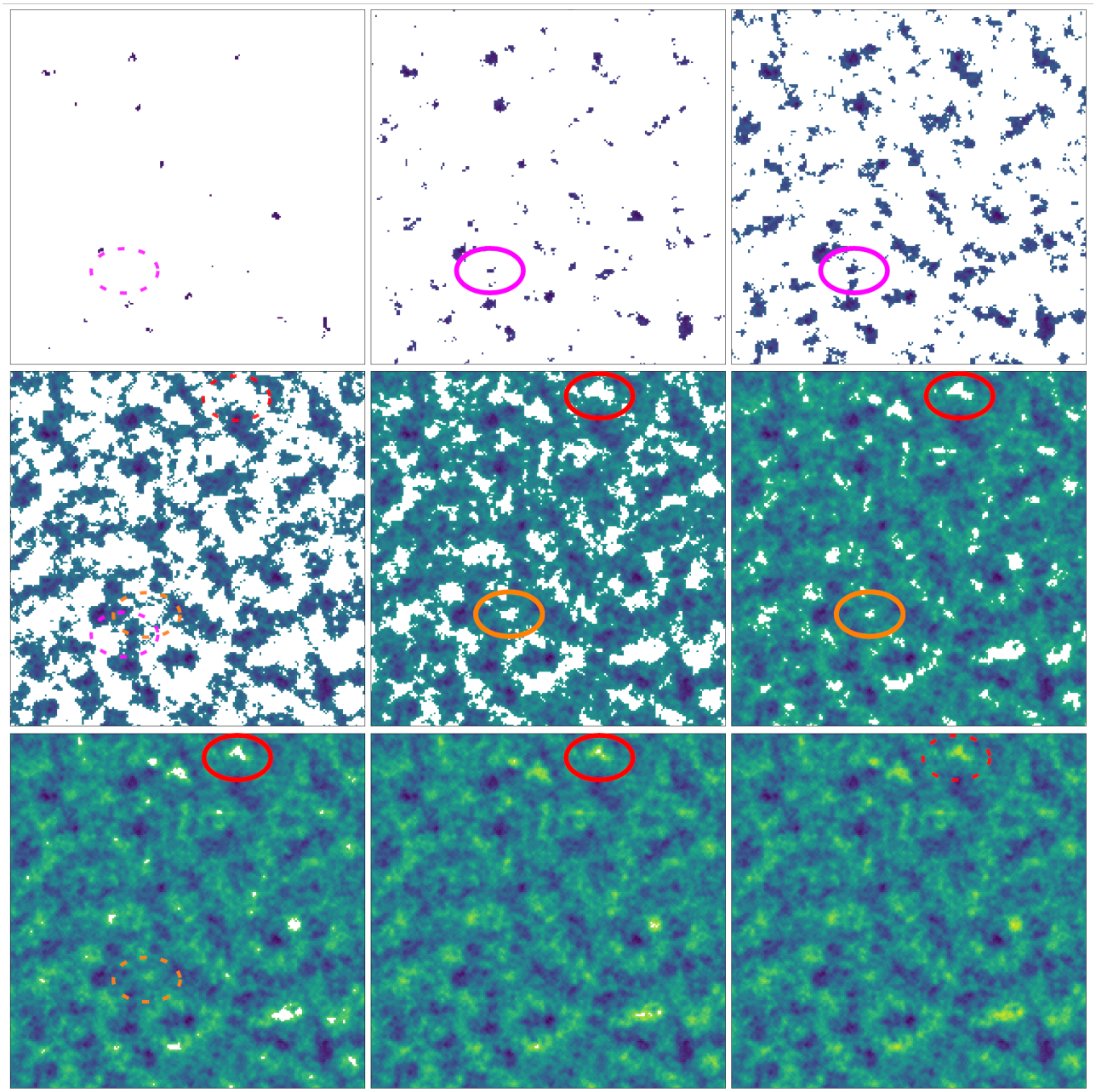


Fig. 2. Excursion sets of a sample signal-to-noise map for a $3.3 \times 3.3\,\mathrm{deg}^2$ field of SLICS for nine values $t_1, \ldots, t_9$. The solid magenta ellipse highlights a connected component corresponding to a minimum of the signal-to-noise map; the solid orange and red ellipses each highlight a hole corresponding to a maximum. The dashed ellipses indicate the position of the features before their birth and after their death. The magenta feature is born at t_2 and dies at t_4 after being absorbed by older features that were born in t_1. This lifeline is described by the interval $[t_2, t_4)$ in $\mathrm{Dgm}(H_0(\mathbb{X}))$. Both the red and orange features are born at t_5. The orange one dies at t_7, while the red one, which corresponds to a sharper peak, persists longer and dies at t_9. Hence, they give rise to intervals $[t_5, t_7)$ and $[t_5, t_9)$ in $\mathrm{Dgm}(H_1(\mathbb{X}))$.

Relative homology is a variant of ordinary homology, which generally speaking can be thought of as an algorithm that takes as an input a space X and a subspace $M \subseteq X$, and gives as an output for each n a vector space $H_n(X, M)$ that describes the features of X that lie outside of M. In our setup, we consider the masked regions as the subspace M of our field X. As the signal-to-noise map takes the constant value $-\infty$ on this subspace, we have $M \subseteq X_i$ for all i and we get a sequence $(H_n(X_i, M))_i$ of relative homology groups that we want to understand.

Just as before, the inclusions $X_i \to X_{i+1}$ make this sequence into a persistence module $H_n(\mathbb{X}, M)$. In analogy to Eq. (10), its Betti numbers are defined via:

$$\beta_n^M \mathbb{X}(t, t') = \#\left\{[t_i, t_j) \in \mathrm{Dgm}(H_n(\mathbb{X}, M)) \mid t_i \le t \le t' < t_j\right\}. \quad (12)$$

Although the definition of these functions uses relative homology, they can be computed using only ordinary homology: From the long exact sequence for relative homology (Hatcher 2002, p. 117), one can deduce the following formula, which expresses these numbers in terms of the Betti numbers of $\mathbb{X}$ and of M:

$$\beta_n^M \mathbb{X}(t_i, t_j) = \beta_n \mathbb{X}(t_i, t_j) - \beta_n \mathbb{X}(-\infty, t_j) + \beta_{n-1}(M) - \beta_{n-1}\mathbb{X}(-\infty, t_i). \quad (13)$$

In our case, where X is a proper subspace of the two-sphere S^2, the only non-zero terms of these invariants are:

$$\begin{aligned}
\beta_0^M\mathbb{X}(t_i,t_j) &= \beta_0\mathbb{X}(t_i,t_j) - \beta_0\mathbb{X}(-\infty,t_j),\\
\beta_1^M\mathbb{X}(t_i,t_j) &= \beta_1\mathbb{X}(t_i,t_j) - \beta_1\mathbb{X}(-\infty,t_j) + \beta_0(M) - \beta_0\mathbb{X}(-\infty,t_i),\\
\beta_2^M\mathbb{X}(t_i,t_j) &= \beta_1(M) - \beta_1\mathbb{X}(-\infty,t_i). \qquad (14)
\end{aligned}$$

These are the quantities that we want to compute, and the next section details how this is numerically done.

2.3.4. Computing persistent Homology: Grids and complexes

As explained in Sect. 2.1, we computed the signal-to-noise maps on a grid with square-shaped pixels. From a mathematical perspective, subdividing X in this way gives it the structure of a finite cubical complex, meaning a space that is obtained by glueing together cubes of maximal dimension (in this case squares) along lower-dimensional cubes (in this case edges and vertices). We obtain a function on this complex by assigning to each square the value of the map on the corresponding pixel and extending this using the lower-star filtration, meaning that an edge or a vertex gets assigned the minimal filtration value of a square it is adjacent to. By construction, the map takes a finite number of values $t_1 \leq \ldots \leq t_k$ on this complex, which define the filtration $\mathbb{X}$. We compute the persistent Betti numbers of this filtered complex using the CUBICAL COMPLEXES module of GUDHI (Dlotko 2020)[12]. The masked regions form a subcomplex M of X and we compute $\beta_n^M\mathbb{X}(t,t')$ using Eq. (13). The number k of steps in the filtration we obtain is quite large, of the order of the number of pixels. GUDHI computes the persistent homology using all these filtration steps, however we evaluate the persistent Betti functions only at a few values (see Appendix B).

With X the full KV450-like footprint, the full corresponding complex is assembled from 17 tiles (see Sect. 2.1), which we label $T^1,\ldots,T^{17}$. In the true KV450 survey, these tiles lie adjacent to each other, and large-scale structures extend across the boundaries. In the simulations however, each tile is constructed from a semi-independent realisation, with no tile-to-tile correlation. Therefore, we perform the calculation of aperture masses and the extraction of Betti numbers for each tile individually, making sure that the boundary of each tile is contained in the mask. This implies that the Betti functions $\beta_n^M\mathbb{X}(t,t')$ of the entire footprint can be computed as the sum

$$\beta_n^M\mathbb{X}(t,t') = \sum_{i=1}^{17}\beta_n^M\mathbb{T}^i(t,t'). \qquad (15)$$

The reason for this is that in X, any two points lying in different tiles are separated by points in the mask. Intuitively speaking, this implies that any feature (e.g., a connected component or a hole) of X that lies outside the masked regions is entirely contained in one of the tiles T^i. Hence, counting the features in X is the same as counting them in each tile and summing them up[13].

2.4. Predicting Betti numbers

As mentioned before, Betti numbers are a highly non-linear statistic, and we were unable to theoretically predict them for a given set of cosmological parameters. Due to the high computational cost of N-body simulations, it would also be completely impossible to perform a set of simulations for every point in our given parameter space. Therefore, we chose to emulate the Betti functions for a chosen set of filtration values by computing them in each simulation of cosmo-SLICS, and interpolating the results using the GAUSSIAN PROCESS REGRESSION (GP regression) of scikit-learn (Pedregosa et al. 2011)[14].

Simply put, a GP regressor is a machine learning algorithm that takes a training data set $\mathcal{D}$, which consists of n observations of a d-dimensional data vector $\boldsymbol{y}_i$, along with errors in the training data set $\sigma^2(\boldsymbol{y})$ and a list of the respective training nodes $\boldsymbol{\pi}$. After training, the emulator provides predictions $\boldsymbol{y}^*$ and their uncertainties $\sigma^2(\boldsymbol{y}^*)$ for arbitrary coordinates $\boldsymbol{\pi}^*$. In our case, the data vectors $\boldsymbol{y}_i$ are the Betti functions extracted from the $n = 26$ different cosmo-SLICS cosmologies. We set the errors in the training data set $\sigma^2(\boldsymbol{y})$ as the variance measured between the different realisations of the light-cone, which varies with cosmology. Our training nodes $\boldsymbol{\pi}$ are the sets of cosmological parameters $\{\Omega_\mathrm{m}, \sigma_8, h, w_0\}$ in each cosmo-SLICS, listed in Table A.1. This method offers a model-independent, probabilistic interpolation of multi-dimensional data sets.

As our kernel, we chose the anisotropic Radial-basis function

$$k(\boldsymbol{\pi}_i,\boldsymbol{\pi}_j) = A\,\exp\left(-\frac{1}{2}d(\boldsymbol{\pi}_i/\boldsymbol{l},\boldsymbol{\pi}_j/\boldsymbol{l})^2\right), \qquad (16)$$

where $\boldsymbol{l}$ is a vector with the same number of dimensions as the input values $\boldsymbol{\pi}_i$ (in this case, division is defined element-wise) and A is a scalar. This kernel determines how similar two points $\boldsymbol{\pi}_i, \boldsymbol{\pi}_j$ are to each other in order to then determine the weights for the individual data points in the interpolation. For each filtration value, we then determined the best hyper-parameters $(A, \boldsymbol{l})$ by minimising the log-marginal-likelihood using a gradient descent from 400 different, randomly chosen initial values. For a detailed description of GP regression we refer the interested reader to Appendix A of H+19 and references therein.

2.5. Peak count statistics

As mentioned in the introduction, we assessed the performance of our topological data analysis by comparing its constraining power to that of the peak count statistics, which is another powerful method to capture information of the non-Gaussian part of the matter distribution. This statistic has been increasingly studied in the literature (e.g., M+18; Kacprzak et al. 2016; Shan et al. 2018; Liu et al. 2015a,b) and is relatively straightforward: It identifies and counts the maxima in the S/N maps, and bins the results as a function of the S/N value at the peak. Peaks of large S/N values convey the majority of the cosmological information (M+18) and correspond to the holes that appear at the latest stages in the filtration sequence presented before. They are typically associated with large galaxy clusters and are less affected by shape noise than peaks of lower S/N values, which however capture additional information from the large-scale structures. In addition to being easy to implement, their constraining power surpasses that of Minkowski functionals (Zürcher et al. 2021) and is competitive with two-point statistics (M+18), making them ideally suited for a performance comparison.

For consistency with the Betti function analysis, we ran our peak finder on the exact same S/N maps. For every simulated

[12] GUDHI is a well-established, open-source program (available in C++ with a python-interface) for topological data analysis. Among many other applications, it can calculate persistent Betti numbers of multi-dimensional fields. The program is publicly available at https://github.com/GUDHI/

[13] More formally, we have $H_n(X,M) \cong \tilde{H}_n(X/M)$ and this quotient space is given by the wedge sum $X/M \cong \bigvee_{i=1}^{17} T^i/M^i$.

[14] Our algorithm was inspired by the GPR EMULATOR tool by Benjamin Giblin (https://github.com/benjamingiblin/GPR_Emulator).

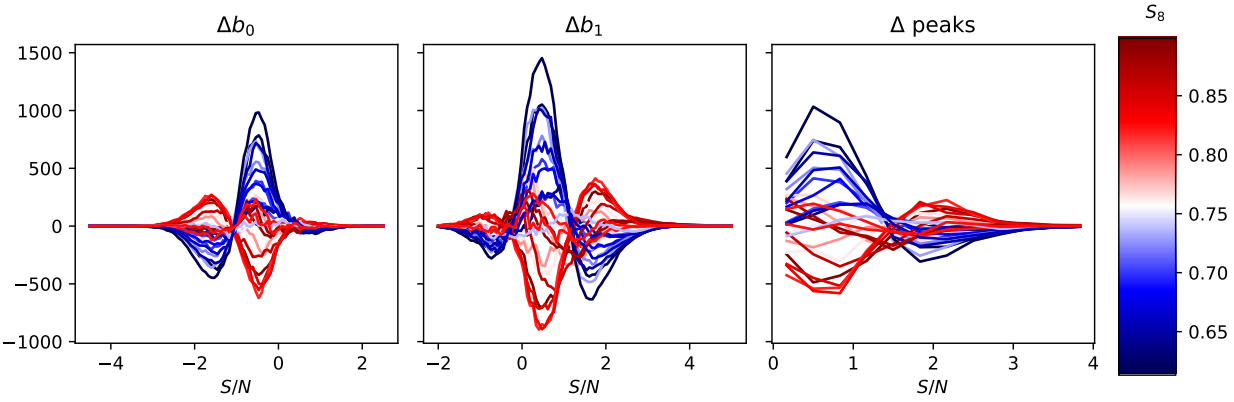


Fig. 3. Difference of average Betti numbers b_0 (*left*) and b_1 (*centre*) and peak counts (*right*) between cosmo-SLICS and SLICS. *Left* and *central panels*: x-axis represents the respective filtration level t, which runs from the minimum to the maximum value of the S/N aperture map. The lines are colour-coded by the value of S_8 in the respective cosmo-SLICS (see Table A.1). The lines shown here are averages over all mock realisations of the full survey footprint (10 for each cosmo-SLICS node, 126 for SLICS).

realisation of the KV450 survey, we counted the peaks on the 17 individual tiles and added the results afterwards. Following M+18, we binned the results in 12 S/N bins ranging from 0.0 to 4.0 and ignored the peaks outside this range. In the rightmost panel of Fig. 3, we present the peak distribution relative to the mean measurement from the SLICS, colour-coded with the S_8 value from the input cosmo-SLICS cosmology. The strong colour gradient illustrates the significant dependence of the signal on S_8.

As explained above, Betti functions are related to the numbers of peaks as well. However, the two methods do not yield completely equivalent information. First, Betti functions do not only take into account the local maxima, but also the minima. However, this has only a small effect on cosmological parameter constraints since most of the information resides in the peaks (M+18, see also Appendix B). Second, and more importantly, peak count statistics are very local in the sense that they decide about whether a pixel is counted as a peak by simply comparing it to the eight adjacent pixels.

Figure 4 illustrates this with a toy example, presenting three simple cases of reconstructed maps. The first two maps cannot be distinguished by peak counts as they both have four peaks with height 1, 2, 3, and 4. In contrast, Betti functions are able to identify structures at larger scales such as regions with high S/N values, and are able to differentiate between the first two panels in Fig. 4: In the first map, the excursion set X_0 has four holes and these get filled one by one as the threshold t increases. We have

$$\beta_1(X_t) = \begin{cases} 4, & 0 \le t < 1, \\ 3, & 1 \le t < 2, \\ 2, & 2 \le t < 3, \\ 1, & 3 \le t < 4, \\ 0, & 4 \le t. \end{cases} \tag{17}$$

This is different from the second map, where for all $0 \le t < 3$, the excursion set X_t has only two holes (although the positions of the holes change),

$$\beta_1(X_t) = \begin{cases} 2, & 0 \le t < 3, \\ 1, & 3 \le t < 4, \\ 0, & 4 \le t. \end{cases} \tag{18}$$

However, while the peaks can differentiate between the second and the third map, the sequence of Betti numbers are the same. In the third map as well, we have two holes for all t between 0 and 3 and these get filled at $t = 3$ and $t = 4$. This illustrates that non-persistent Betti functions cannot distinguish very well between a small number of sharp peaks and a high number of peaks that only slightly protrude from their surroundings. (We ignore β_0 here as for all $t \ge 0$, all excursion sets have exactly one connected component). Persistent Betti functions can differentiate all three cases. They are able to detect that in the second map, the positions of the two holes at $0 \le t < 3$ vary, while in the third map, the same two holes just get filled while t increases. The persistence diagrams $\mathrm{Dgm}(H_1(\mathbb{X}))$ associated with the three panels are:

$$\{[0,4),[0,3),[0,2),[0,1)\}, \quad \{[0,4),[0,1),[1,2),[2,3)\}, \quad \text{and } \{[0,4),[0,3)\}. \tag{19}$$

As all these are different, so are the associated persistent Betti functions $\beta_1\mathbb{X}$. This simple example illustrates how persistent Betti numbers combine the local information from peak count statistics with the non-local, large-scale information from non-persistent Betti functions.

3. Results

In this section, we present the results from our topological analysis based on the persistent Betti numbers. We gauge the performance of the technique from a comparison with an analysis based on their non-persistent alternatives, and additionally with a peak statistics analysis. We first present our measurements from the simulations, which serve to train the emulator and estimate the covariance matrices, then proceed with the parameter inference.

3.1. Calibrating the emulator and determining the covariances

For each light-cone of the cosmo-SLICS simulation suite, we measured the persistent Betti functions $b_n^{\mathrm{per}}(t,t') = \beta_n^M\mathbb{X}(t,t')$, and computed a non-persistent version b_n by setting $b_n(t) = b_n^{\mathrm{per}}(t,t) = \beta_n^M\mathbb{X}(t,t)$. We extracted the average values of $b_{0,1}$ and $b_{0,1}^{\mathrm{per}}$ at a chosen set of values for each of the 26 different cosmologies from the mean over the ten light-cones.

To demonstrate that Betti functions are indeed sensitive to the underlying cosmological parameters, we show the

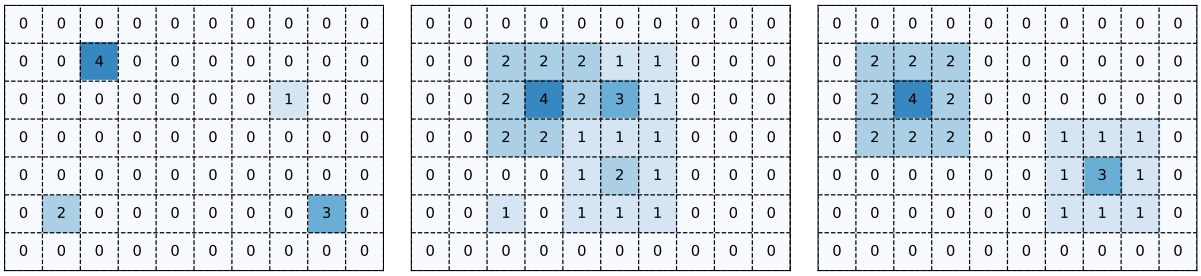


Fig. 4. Three idealised S/N maps. Both peak count statistics and non-persistent Betti functions can distinguish the first from the third map. However, the peak counts of the first and the second map are identical, as are the non-persistent Betti functions of the second and the third. Persistent Betti functions can distinguish all three maps. For further explanations, see Sect. 2.5.

non-persistent Betti functions[15] b_n for different input cosmologies in Fig. 3, again colour-coded as a function of S_8. As for the peak statistics, there is a strong colour gradient at every element of the data vector, both for b_0 and b_1, which indicates the sensitivity to that parameter.

In both the non-persistent and persistent settings, we obtained a data vector consisting of the function values at the chosen evaluation points (introduced in Sect. 2.3.4 and discussed in Appendix B), and further concatenated these values for b_0 and b_1 or $b_0^{\rm per}$ and $b_1^{\rm per}$, respectively. We did not consider b_2 and $b_2^{\rm per}$ for our analysis as they only consist of a very small number of features, which mainly depend on the shape of the mask and contain very little cosmological information. We used these data vectors to train the hyperparameters of our GPR emulator, and were hereafter able to predict the Betti functions (at the chosen evaluation points) for arbitrary cosmological parameters enclosed within the training range. As a next step, we extracted the same Betti functions for each realisation of the SLICS and used these to determine a covariance matrix $C_{\rm b}$ for the data vector.

To assess the accuracy of the emulator, we performed a 'leave-one-out cross-validation test': We excluded one cosmology node and trained our emulator on the 25 remaining ones. We then let the emulator predict Betti functions for the previously excluded cosmology and compared it to the measured values; this process was repeated for each of the 26 cosmologies. The results are presented in Appendix C. As further discussed therein, this test demonstrates that we achieve an accuracy of a few percent on our predicted signal at most evaluation points. While a single outlier contains inaccuracies that are larger than the 1σ limit of the sample variance extracted from SLICS, the vast majority of cross-validation tests is contained within this limit.

For each set of cosmological parameters, the emulator not only provides a prediction for the Betti functions, but also an error estimate. We define the 'covariance matrix' of the emulator $C_{\rm e}$ as a diagonal matrix where the entries correspond to the uncertainty that is predicted by the GPR emulator. Combining with the sample variance $C_{\rm b}$ estimated from the SLICS, we subsequently set

$$C = C_{\rm b} + C_{\rm e} \tag{20}$$

as the covariance matrix for our cosmological parameter estimation. The emulator covariance $C_{\rm e}$ depends on the chosen cosmological parameters, therefore our covariance matrix C varies with cosmology. We note that it would also be possible to obtain an error estimate $C_{\rm e}$ from the cross-validation: For each evaluation point of the Betti functions, one can estimate the variance by

$$\sigma^2_{b_n} = \frac{1}{25}\sum_{i=0}^{25}\left(x_{i,\rm measured} - x_{i,\rm predicted}\right)^2, \tag{21}$$

where $x_{i,\rm measured}$ is the measured value of the Betti function at the ith training cosmology, whereas $x_{i,\rm predicted}$ is the one predicted by the emulator when it is trained on all other cosmologies, leaving the ith one out. We verified that both the error estimates themselves and the posteriors on the cosmological parameters are consistent between the two methods.

We finally repeated this full machinery (GPR training and covariance estimation), this time on the peak count statistics measurements extracted from the same data sets and described in Sect. 2.5. We compare the respective performances in the next section.

3.2. Inference of cosmological parameters

Putting to use the covariance matrix and the prediction tools described in the last section, we performed a trial cosmological parameter inference using the mean of the SLICS measurements as our 'observed data' to test how well we can recover the fiducial values, and how the constraining power on the cosmological parameters compares with the peak count statistics described in Sect. 2.5. We sampled our cosmological parameter space using the MCMC sampler EMCEE (Foreman-Mackey et al. 2013), specifying a flat prior range that reflects the parameter space sampled by the cosmo-SLICS, namely: $\Omega_{\rm m} \in [0.1, 0.6]$, $S_8 \in [0.6, 1.0]$, $h \in [0.5, 0.9]$, $w_0 \in [-1.8, -0.2]$. Since our covariance matrix is extracted from a finite number of simulations, it is inherently noisy, which leads to a biased inverse covariance matrix (Hartlap et al. 2007). To mitigate this, we adopted a multivariate t-distribution (Sellentin & Heavens 2016) as our likelihood model. At each sampled point, we computed the log-likelihood as

$$\log(\mathcal{L}) = -\frac{1}{2}N_{\rm s}\log\left(1 + \frac{\chi^2}{N_{\rm s}-1}\right), \tag{22}$$

where $N_{\rm s}$ is the number of simulations used to calibrate the covariance matrix (in our case 126), and χ^2 was computed in the usual way via

$$\chi^2 = (\boldsymbol{x}_{\rm measured} - \boldsymbol{x}_{\rm predicted})^{\rm T}\, C^{-1}\, (\boldsymbol{x}_{\rm measured} - \boldsymbol{x}_{\rm predicted}). \tag{23}$$

Here, $\boldsymbol{x}_{\rm measured}$ are the respective Betti functions $b_{0,1}$ or $b_{0,1}^{\rm per}$ extracted from SLICS, while $\boldsymbol{x}_{\rm predicted}$ are predicted by the emulator at the respective point in parameter space[16].

[15] We focus on the non-persistent case here, since the persistent Betti numbers depend on tuples of evaluation points, which cannot be visualised easily.

[16] We note that this model assumes a Gaussian likelihood for the persistent Betti numbers; we have verified that our data vectors extracted from SLICS follow an approximately Gaussian distribution and leave a rigorous quantification of Gaussianity for future work.

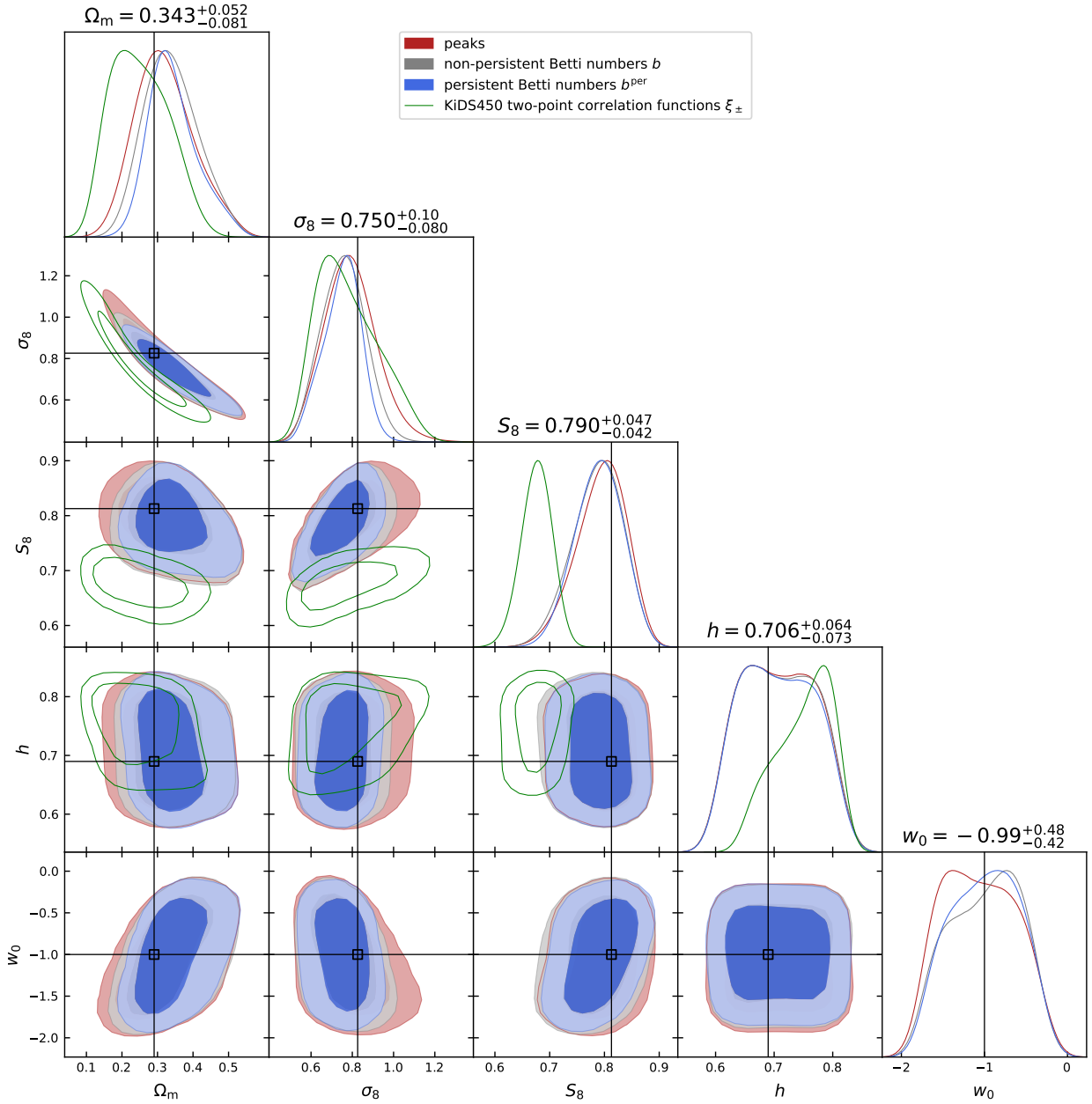


Fig. 5. Results of an MCMC on the Betti numbers measured in the KV450-like SLICS. The contours represent the 1- and 2-σ posterior contours for an analysis with non-persistent (grey) and persistent (blue) Betti functions. For comparison, the red contours represent the posterior of an analysis with peak statistics, as done in M+18. The values on top of the diagonals represent the marginalised posterior for the persistent Betti functions b^{per}. The fiducial parameters of SLICS are represented by the black lines. We conducted our analysis with Ω_m, σ_8, h and w_0 as free parameters and employed flat priors with $\Omega_m \in [0.1, 0.6]$, $S_8 \in [0.6, 1.0]$, $h \in [0.5, 0.9]$, $w_0 \in [-1.8, -0.2]$. The green lines correspond to the 1- and 2-σ posterior contours of a cosmological parameter analysis of the KiDS450-survey using tomographic two-point correlation functions, where the covariance matrix was extracted from SLICS and systematics were disregarded (see the 'N-body' setup in Table 4 of Hildebrandt et al. 2017). Since the setup of that analysis is very similar to ours, they can be used to compare the relative sizes (but not locations) of contours. We note that this parameter analysis was done in a ΛCDM-framework on KiDS data, whereas our analysis is done in a wCDM framework on mock data.

Following these steps, we performed an MCMC analysis both for the persistent and the non-persistent Betti functions, $b^{per}_{0,1}$ and $b_{0,1}$ and for the peak count statistics. The results are displayed in Fig. 5. As can be seen, all methods recover the fiducial simulation parameter within the 1σ limit. Furthermore, the persistent Betti functions b^{per} offer slightly improved constraining power when compared to the non-persistent version b and to peak counts, and is notably the best at rejecting the low-Ω_m and the high-σ_8 regions of the parameter space. The 1σ contours are all smaller, yielding marginalised constraints on S_8 of $0.790^{+0.047}_{-0.042}$, compared to $0.788^{+0.051}_{-0.041}$ for non-persistent Betti numbers and $0.794^{+0.053}_{-0.039}$ for peaks. Persistent homology therefore increases the constraining power on S_8 by 3% compared to peak statistics, while the 'figure of merit' (i.e. one over the area of the 1σ contour) in the $\Omega_m - \sigma_8$ plane is increased by 48%. This is not entirely surprising as b and peaks just counts features at each filtration step, whereas b^{per} contains the entire information content of the persistence module, as discussed in Sect. 2.3.2.

For comparison, we also report in Fig. 5 the constraints from a two-point correlation function analysis (see the green curves).

These were obtained from the actual measurements from the KiDS-450 data, which had a slightly different redshift distribution and number density than the KV450-like data used in this paper, but were derived with a covariance matrix extracted from the SLICS and ignored systematic uncertainty (see the N-body setup in Table 4 of Hildebrandt et al. 2017), in a setup otherwise very similar to ours. The contours are offset to lower S_8 values as preferred by the KiDS-450 data, in contrast with the higher input SLICS cosmology. Due to the very similar analysis setup, the contours can be used to compare their size with the ones derived from our analysis. The increased constraining power with respect to peaks and homology is largely due to the tomographic analysis, which we reserve for future work.

It is worth noting that a part of the reported uncertainties for persistent Betti functions arise due to the inaccuracy of the emulator that is larger than for peaks; when these uncertainties are ignored, persistent Betti numbers achieve constraints of $S_8 = 0.787^{+0.045}_{-0.040}$, an improvement of 8% with respect to peak statistics. This can potentially be realised in a future analysis involving a larger training set (we further discuss this in Appendix C). Although a gain of 3−8% in accuracy on S_8 is small in comparison with the infrastructure work required, we show in the next section that the improvement is significantly increased once the weak lensing data better resolves the large-scale structure around the peaks.

Before steering away from the KV450-like setup however, we finally investigate the systematic impact of two other survey parameters. Similarly to the shear peak count statistics, the Betti functions are extracted from the S/N of the aperture mass map. Therefore, their expectation value directly depends not only on the survey footprint but also on various terms influencing the noise, for example the effective number density of galaxies and their average internal ellipticity (see e.g., M+18). For the two-point correlation functions, these terms only influence the covariance matrix, at least to first order (Heydenreich et al. 2020), while here a 10% offset on the galaxy density or on the shape noise can bias the inferred parameters by $1-3\sigma$. Full details of these tests are provided in Appendix D.

4. Outlook for Stage IV experiments

As discussed in Sect. 2.5, the gain of information from persistent Betti functions with respect to peak counting comes from the ability to extract information about the environment, for example by distinguishing whether peaks are isolated or clustered on a larger over-density. Consequently, we expect that resolving the large-scale structure with higher resolution, as will be made possible with upcoming Stage IV lensing experiments, should further increase this relative gain.

In this section, we explore this scenario by repeating the analysis presented previously on a set of simulations with *Euclid*-like source number densities. These are also constructed from the SLICS and cosmo-SLICS light-cones, but differ in a few key aspects: In contrast to the KV450-like mock data, the position of the galaxies are here placed at random on the $10 \times 10\,\mathrm{deg}^2$ fields, and no masks are imposed on them. Our total survey area is therefore $100\,\mathrm{deg}^2$. The redshift distribution follows:

$$n(z) \propto z^2 \exp\left[-\left(\frac{z}{z_0}\right)^{\beta}\right], \tag{24}$$

with $z_0 = 0.637, \beta = 1.5$, and the overall proportionality constant is given by normalising the distribution to $30\,\mathrm{gal\,arcmin}^{-2}$.

In this analysis, we opted for an aggressive strategy in which we include peaks and features with S/N up to ten, which might end up being rejected in the future due to difficulties at modelling all systematic effects (e.g., baryons feedback) at the required level. We carried out a second analysis which cap the features at S/N of seven instead, and noticed only minor differences. In both cases, the peak statistics were binned such that a signal-to-noise range of one is covered by three bins.

One important aspect to note is that due to the much lower level of shape-noise in this Stage IV setup, the accuracy of the GPR emulator degrades in comparison to the KV450-like scenario, reaching ∼5% only for a few evaluation points. More details about the relative importance of the emulator accuracy are provided in Appendix C. Meanwhile, we have included the error associated with the emulator following Eqs. (20) and (21) in our analysis, hence our results should remain unbiased.

The results of our MCMC analysis are shown in Fig. 6, where we compare the performance of the persistent Betti numbers to peak counting. We find here again an increase in statistical power and the gain is amplified compared with the KV450-like analysis: Looking at the S_8 constraints, we measure $S_8 = 0.796^{+0.032}_{-0.032}$ for peak statistics and $S_8 = 0.795^{+0.028}_{-0.024}$ for persistent Betti numbers, an improvement of ∼19%. Contraints on Ω_m are also improved, with a gain of ∼12% on the one-dimensional marginal error. Persistent Betti numbers are even able to set some constraints on the dark energy equation of state, measuring $w_0 = -0.90^{+0.33}_{-0.22}$. This is an exciting new avenue that will be further investigated in future work involving tomographic decomposition of the source catalogues.

We emphasise that for the persistent Betti numbers, the emulator inaccuracy constitutes a part of the error budget. This uncertainty could be reduced by increasing the number of training nodes (see Appendix C). When disregarding the emulator uncertainty, we achieve constraints of $S_8 = 0.795^{+0.027}_{-0.023}$ and $w_0 = -0.91^{+0.27}_{-0.20}$.

5. Discussion

Based on the topological analysis of an ensemble of realistic numerical simulations, we have demonstrated that persistent Betti functions are a highly competitive method to constrain cosmological parameters from weak lensing data. We carried out an MCMC sampling and found that persistent Betti functions improve the constraining power on S_8 by ∼3−8% compared to peak count statistics, which themselves have a constraining power that is similar to the mainstream two-point statistics (M+18). Furthermore, this advantage over peak statistics is expected to exceed 20% in the upcoming Stage IV surveys. We included the effect of shape noise and of masking as they occur within the KV450 weak lensing survey, however, our methods can be directly adapted to other surveys. We believe that this gain in information over peak statistics is primarily caused by the sensitivity of Betti numbers to the large-scale structure. It would be interesting to see whether this sensitivity can be replicated using peak statistics when a set of different filter radii θ_ap is used. However, we believe that persistent Betti numbers offer a simpler and numerically less expensive alternative to that.

Our results are obtained by estimating a covariance matrix from an ensemble of independent light cones, and by training a Gaussian Process Regression emulator on a suite of wCDM simulations. A non-negligible part of our error budget arises from

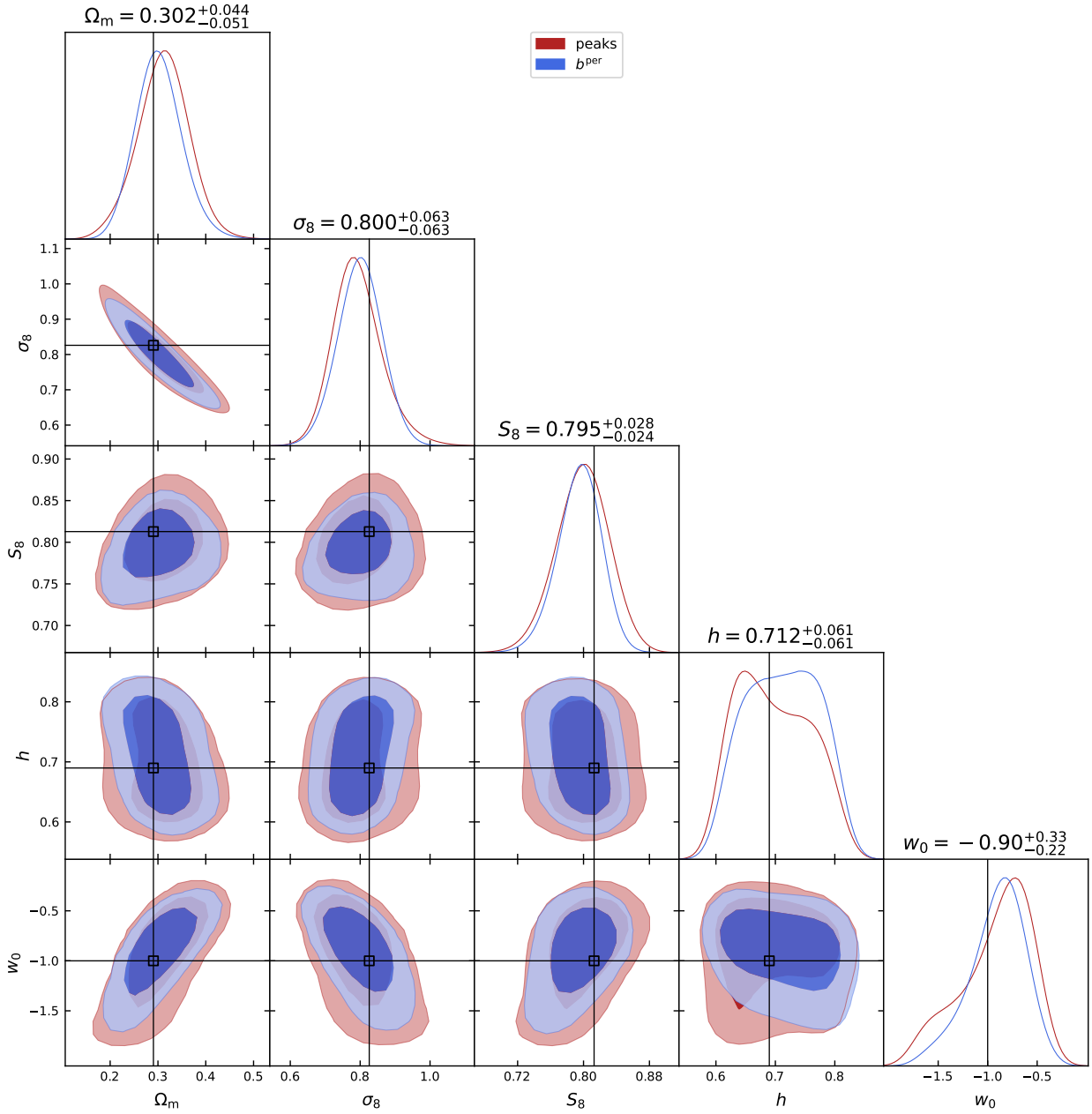


Fig. 6. Results of an MCMC as described in Fig. 5, but here for 100 deg^2 of *Euclid*-like lensing data.

inaccuracies in the emulator. These occur since we can only calibrate our model for 26 different cosmologies, which can certainly be improved in the future.

We investigated the influence of shape noise and number density on our analysis (see Appendix D). However, various other systematic effects inherent to weak lensing data analyses were not explored in this paper, including the uncertainty arising from photometric redshift errors, shape calibration, baryon feedback on the matter distribution or intrinsic alignments of galaxies. Before a full cosmological parameter analysis can be carried out from data, it is crucial to investigate and understand how to include these. Additionally, there are many internal parameters in our analysis pipeline that were chosen by hand, for example the aperture radius θ_{ap} of the filter function or the points at which we evaluate the Betti functions. With careful optimisation, one would probably be able to extract an even higher amount of information from the persistent Betti numbers. In future work, it would be interesting to see if forward modelling techniques such as DELFI (Taylor et al. 2019) or BORG (Jasche & Lavaux 2019) can also be used in combination with Betti functions.

We also want to point out that persistent Betti numbers are just one way to compress the information of persistent homology. While they are certainly easy to understand and apply, they also suffer from some disadvantages, the most notable of which being that the difference between two Betti functions is always integer-valued, whereas we would prefer a real-valued distance function for a χ^2-analysis. This problem can be mitigated by utilising different statistics of persistent homology (e.g., Reininghaus et al. 2015; Bubenik 2015); for an overview of further options, see also Oudot (2015, Chap. 8) and Pun et al. (2018).

Lastly, we note that topological data analysis is a promising avenue in cosmology that can find multiple applications well outside those presented in this paper. Other methods similar

or complementary to persistent Betti functions can be used to detect and quantify structure in both continuous fields and point-clouds in arbitrary dimensions (see e.g., Otter et al. 2017, and references therein), which makes topological data analysis an incredibly versatile tool to study the distribution of matter in our Universe. While it has been used in cosmology before, for example to detect non-Gaussianities in the CMB (see P+19) or to find voids in the large-scale structure (Xu et al. 2019), its utilisation in modern astronomy is bound to grow in the near future.

Acknowledgements. We are grateful for useful hints from Fabian Lenzen and Vidit Nanda on persistent homology and from Pierre Burger and Peter Schneider on the data analysis. Furthermore we thank Shahab Joudaki, Konrad Kuijken, Frank Röttger, Sandra Unruh and the anonymous referee for helpful comments on previous versions of this article. Sven Heydenreich acknowledges support from the German Research Foundation (DFG SCHN 342/13), the International Max-Planck Research School (IMPRS) and the German Academic Scholarship Foundation. Benjamin Brück was supported by the Danish National Research Foundation through the Copenhagen Centre for Geometry and Topology (DNRF151). Joachim Harnois-Déraps acknowledges support from an STFC Ernest Rutherford Fellowship (project reference ST/S004858/1). Computations for the *N*-body simulations were enabled by Compute Ontario (www.computeontario.ca), Westgrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). The SLICS numerical simulations can be found at http://slics.roe.ac.uk/, while the cosmo-SLICS can be made available upon request. *Author contributions.* All authors contributed to the development and writing of this paper. SH led the data analysis, BB developed the necessary mathematical background, while JHD provided the suites of numerical simulation tailored to the measurement.

References

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2020, Phys. Rev. D, 102
Adler, R. J. 1981, The Geometry of Random Fields (Chichester: John Wiley & Sons, Ltd.)
Angulo, R. E., Zennaro, M., Contreras, S., et al. 2020, ArXiv e-prints [arXiv:2004.06245]
Asgari, M., Tröster, T., Heymans, C., et al. 2020, A&A, 634, A127
Bartelmann, M., & Schneider, P. 2001, Phys. Rep., 340, 291
Blaser, N., & Brun, M. 2019, ArXiv e-prints [arXiv:1911.07484]
Bresten, C., & Jung, J. H. 2019, ArXiv e-prints [arXiv:1910.08245]
Bubenik, P. 2015, J. Mach. Learn. Res., 16, 77
Burger, P., Schneider, P., Demchenko, V., et al. 2020, A&A, 642, A161
Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, MNRAS, 499, 5902
de Silva, V., Morozov, D., & Vejdemo-Johansson, M. 2011, Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data, 27, 124003
Dlotko, P. 2020, GUDHI User and Reference Manual, 3.1.1 Edition (GUDHI Editorial Board)
Edelsbrunner, H., & Harer, J. 2008, in Surveys on Discrete and Computational Geometry (Providence, RI: American Mathematical Society), Contemp. Math., 453, 257
Elbers, W., & van de Weygaert, R. 2019, MNRAS, 486, 1523
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fu, L., Kilbinger, M., Erben, T., et al. 2014, MNRAS, 441, 2725
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2004, Bayesian Data Analysis, 2nd edn. (Chapman and Hall/CRC)
Ghrist, R. 2014, Elementary Applied Topology, 1st edn. (Createspace)
Giblin, B., Heymans, C., Harnois-Déraps, J., et al. 2018, MNRAS, 480, 5529
Gruen, D., Friedrich, O., Krause, E., et al. 2018, Phys. Rev. D, 98
Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, PASJ, 72, 16
Hamilton, A. J. S., Gott, J. R., III, & Weinberg, D. 1986, ApJ, 309, 1
Harnois-Déraps, J., & van Waerbeke, L. 2015, MNRAS, 450, 2857
Harnois-Déraps, J., Amon, A., Choi, A., et al. 2018, MNRAS, 481, 1337
Harnois-Déraps, J., Giblin, B., & Joachimi, B. 2019, A&A, 631, A160
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Hatcher, A. 2002, Algebraic Topology (Cambridge: Cambridge University Press)
Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. 2014, ApJ, 780, 111
Hetterscheidt, M., Erben, T., Schneider, P., et al. 2005, A&A, 442, 43
Heydenreich, S., Schneider, P., Hildebrandt, H., et al. 2020, A&A, 634, A104
Heymans, C., Grocutt, E., Heavens, A., et al. 2013, MNRAS, 432, 2433
Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Ivezic, Z., Axelrod, T., Brandt, W. N., et al. 2008, Serb. Astron. J., 176, 1
Jasche, J., & Lavaux, G. 2019, A&A, 625, A64
Jeffrey, N., Abdalla, F. B., Lahav, O., et al. 2018, MNRAS, 479, 2871
Joudaki, S., Blake, C., Heymans, C., et al. 2017, MNRAS, 465, 2033
Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, A&A, 638, L1
Kacprzak, T., Kirk, D., Friedrich, O., et al. 2016, MNRAS, 463, 3653
Kaiser, N. 1992, ApJ, 388, 272
Kaiser, N., & Squires, G. 1993, ApJ, 404, 441
Kilbinger, M., Fu, L., Heymans, C., et al. 2013, MNRAS, 430, 2200
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Liu, J., & Madhavacheril, M. S. 2019, Phys. Rev. D, 99
Liu, X., Pan, C., Li, R., et al. 2015a, MNRAS, 450, 2888
Liu, J., Petri, A., Haiman, Z., et al. 2015b, Phys. Rev. D, 91
Lo, D., & Park, B. 2018, PLoS One, 13
MacPherson, R., & Schweinhart, B. 2012, J. Math. Phys., 53
Makarenko, I., Shukurov, A., Henderson, R., et al. 2018, MNRAS, 475, 1843
Marques, G. A., Liu, J., Zorrilla Matilla, J. M., et al. 2019, JCAP, 2019, 019
Martinet, N., Schneider, P., Hildebrandt, H., et al. 2018, MNRAS, 474, 712
Mootoovaloo, A., Heavens, A. F., Jaffe, A. H., & Leclercq, F. 2020, MNRAS, 497, 2213
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Nicolau, M., Levine, A. J., & Carlsson, G. 2011, Proc. Natl. Acad. Sci., 108, 7265
Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. 2017, EPJ Data Sci., 6, A17
Oudot, S. Y. 2015, in Persistence Theory: from Quiver Representations to Data Analysis (Providence, RI: American Mathematical Society), Math. Surv. Monogr., 209, 218
Parroni, C., Cardone, V. F., Maoli, R., & Scaramella, R. 2020, A&A, 633, A71
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Petri, A., Liu, J., Haiman, Z., et al. 2015, Phys. Rev. D, 91
Planck Collaboration VI. 2020, A&A, 641, A6
Pokorny, F. T., Goldberg, K., & Kragic, D. 2016, 2016 IEEE International Conference on Robotics and Automation (ICRA), 16
Pranav, P., Edelsbrunner, H., van de Weygaert, R., et al. 2017, MNRAS, 465, 4281
Pranav, P., Adler, R. J., Buchert, T., et al. 2019a, A&A, 627, A163
Pranav, P., van de Weygaert, R., Vegter, G., et al. 2019b, MNRAS, 485, 4167
Pun, C. S., Xia, K., & Lee, S. X. 2018, ArXiv e-prints [arXiv:1811.00252]
Reininghaus, J., Huber, S., Bauer, U., & Kwitt, R. 2015, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4741
Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85
Schirmer, M., Erben, T., Hetterscheidt, M., & Schneider, P. 2007, A&A, 462, 875
Schneider, P. 1996, MNRAS, 283, 837
Schneider, P., & Lombardi, M. 2003, A&A, 397, 809
Schneider, P., Eifler, T., & Krause, E. 2010, A&A, 520, A116
Seitz, S., & Schneider, P. 2001, A&A, 374, 740
Sellentin, E., & Heavens, A. F. 2016, MNRAS, 456, L132
Shan, H., Liu, X., Hildebrandt, H., et al. 2018, MNRAS, 474, 1116
Sousbie, T. 2011, MNRAS, 414, 350
Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, ArXiv e-prints [arXiv:1305.5422]
Taylor, P. L., Kitching, T. D., Alsing, J., et al. 2019, Phys. Rev. D, 100
Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, Phys. Rev. D, 98
Unruh, S., Schneider, P., Hilbert, S., et al. 2020, A&A, 638, A96
van de Weygaert, R., Vegter, G., Edelsbrunner, H., et al. 2013, ArXiv e-prints [arXiv:1306.3640]
van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, MNRAS, 476, 4662
Wright, A. H., Hildebrandt, H., Kuijken, K., et al. 2019, A&A, 632, A34
Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2020, A&A, 640, L14
Xu, X., Cisewski-Kehe, J., Green, S. B., & Nagai, D. 2019, Astron. Comput., 27, 34
Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., & Refregier, A. 2021, JCAP, 2021, 028

Appendix A: Additional information about the cosmo-SLICS simulations

For completeness, we list in Table A.1 the full suite of cosmological parameters that makes up the cosmo-SLICS series. Full details about this simulation suite can be found in H+19.

Table A.1. Cosmological parameters of the cosmo-SLICS *w*CDM simulations.

	Ω_m	S_8	h	w_0	σ_8	Ω_cdm
00	0.3282	0.6984	0.6766	−1.2376	0.6677	0.2809
01	0.1019	0.7826	0.7104	−1.6154	1.3428	0.0546
02	0.2536	0.6133	0.6238	−1.7698	0.667	0.2063
03	0.1734	0.7284	0.6584	−0.5223	0.9581	0.1261
04	0.3759	0.8986	0.6034	−0.9741	0.8028	0.3286
05	0.4758	0.7618	0.7459	−1.3046	0.6049	0.4285
06	0.1458	0.768	0.8031	−1.4498	1.1017	0.0985
07	0.3099	0.7861	0.694	−1.8784	0.7734	0.2626
08	0.4815	0.6804	0.6374	−0.7737	0.5371	0.4342
09	0.3425	0.7054	0.8006	−1.501	0.6602	0.2952
10	0.5482	0.6375	0.7645	−1.9127	0.4716	0.5009
11	0.2898	0.7218	0.6505	−0.6649	0.7344	0.2425
12	0.4247	0.7511	0.6819	−1.1986	0.6313	0.3774
13	0.3979	0.8476	0.7833	−1.1088	0.736	0.3506
14	0.1691	0.8618	0.789	−1.6903	1.1479	0.1218
15	0.1255	0.6131	0.7567	−0.9878	0.9479	0.0782
16	0.5148	0.8178	0.6691	−1.3812	0.6243	0.4675
17	0.1928	0.8862	0.6285	−0.8564	1.1055	0.1455
18	0.2784	0.65	0.7151	−1.0673	0.6747	0.2311
19	0.2106	0.8759	0.7388	−0.5667	1.0454	0.1633
20	0.443	0.8356	0.6161	−1.7037	0.6876	0.3957
21	0.4062	0.662	0.8129	−1.9866	0.5689	0.3589
22	0.2294	0.8226	0.7706	−0.8602	0.9407	0.1821
23	0.5095	0.7366	0.6988	−0.7164	0.5652	0.4622
24	0.3652	0.6574	0.7271	−1.5414	0.5958	0.3179
Fid	0.2905	0.8231	0.6898	−1.0	0.8364	0.2432

Notes. Other fixed parameters are $\Omega_\mathrm{b} = 0.0447$ and $n_\mathrm{s} = 0.969$.

Appendix B: Choosing evaluation points for the Betti functions

While the domain of both persistent and non-persistent Betti functions is technically limited by the pixel-values of the signal-to-noise map, the raw data vector provided by GUDHI contains about 10^6 entries. As this level of refinement is neither practical nor necessary, we need to choose points at which to evaluate and compare the Betti functions. To help in this decision, we computed the persistent Betti numbers for SLICS and cosmo-SLICS on a dense grid. We then compared the mean squared differences between SLICS and cosmo-SLICS, weighted by the inverse variance of the respective Betti numbers within SLICS. This basically measures how well an evaluation point can distinguish different cosmologies. We rank the evaluation points according to this discrimination potential, starting with the 'best evaluation point. We then recursively built our set of evaluation points in the following way.

Given a set of evaluation points, we extracted the mean of the persistent Betti numbers measured from SLICS, $\boldsymbol{x}_\mathrm{SLICS}$, and the same data vector for each cosmology i of the cosmo-SLICS, $\boldsymbol{x}_i$. Furthermore, we extracted a covariance matrix C for this data vector from SLICS. Using these, we computed the quantity $\tilde{\chi}^2$ via:

$$\tilde{\chi}^2 = \frac{1}{26}\sum_i \left[(\boldsymbol{x}_\mathrm{SLICS} - \boldsymbol{x}_i)^\mathrm{T} C^{-1} (\boldsymbol{x}_\mathrm{SLICS} - \boldsymbol{x}_i)\right]. \tag{B.1}$$

We then added the next evaluation point to our data vector and checked whether the new $\tilde{\chi}^2$ surpassed the old one by a certain threshold (in our case 0.2), while additionally demanding a minimum of 40 features per line-of-sight[17]. If those two criteria were fulfilled, we added this evaluation point to the data vector and repeated the process. If not, we checked the next-best evaluation point. This yields a total of 33 evaluation points for the KV450-like survey, and 46 evaluation points for the *Euclid*-like survey. We note that this method favours b_1 over b_0, consistent with the

Table B.1. Chosen evaluation points for the Betti functions.

Betti function	Evaluation points
b_0	−2.0, −1.8, −1.6, −1.4, −0.7, −0.4, −0.2
b_1	0.1, 0.2, 0.4, 0.6, 0.8, 1.4, 1.6, 1.8, 2.3, 2.6, 3.0, 3.4, 4.0
b_0^per	(−2.0, −1.2), (−2.0, −0.6), (−1.9, −1.8), (−1.7, −1.7), (−1.3, −1.3), (−0.5, 0.0)
b_1^per	(−0.1, 1.1), (0.1, 2.8), (0.1, 3.0), (0.1, 3.1), (0.1, 3.8), (0.1, 3.9), (0.2, 0.2), (0.2, 0.5), (0.7, 2.5), (0.7, 2.6), (0.7, 2.7), (0.7, 4.0), (0.8, 3.9), (0.9, 2.4), (1.0, 2.2), (1.1, 2.1), (1.2, 1.6), (1.4, 1.8), (1.5, 1.7), (1.5, 1.9), (1.5, 2.0), (1.7, 2.6), (2.5, 3.0), (2.6, 4.0), (2.7, 2.8), (3.5, 4.0), (3.8, 4.0)

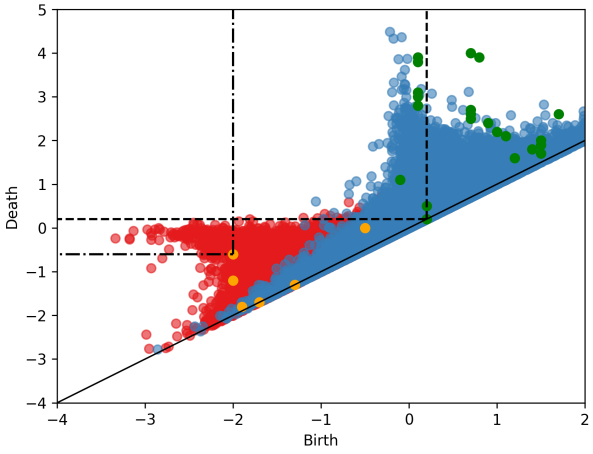


Fig. B.1. Persistence diagram extracted from one tile T^i of the SLICS, in the KV450-like setup. The scattered points represent the features in $\mathrm{Dgm}(H_0(\mathbb{X}))$ (red) and $\mathrm{Dgm}(H_1(\mathbb{X}))$ (blue). For each feature, the x-value corresponds to the S/N-level at which that feature is born, whereas the y-value corresponds to the S/N-level at which it disappears again. The orange and green dots mark our evaluation points of the respective persistent Betti functions $b_{0,1}^\mathrm{per}$. The two dashed rectangles visualise regions in the diagram inside of which all features are counted, when computing the Betti functions at the evaluation points $(-2, -0.6)$ and $(0.2, 0.2)$.

[17] This ensures that we avoid tuning our analysis on artefacts that are dominated by shot-noise.

findings of M+18 that peaks have higher information content than troughs.

To visualise this, we plot all features of the persistence diagram as a scatter plot, where the x-coordinate of a point represents the filtration value at which the respective feature is born and the y-coordinate represents the filtration value of its death. In general, a persistent Betti function evaluated at (x, y) counts all features that lie towards the upper-left of the evaluation point (see Fig. B.1). Here, features that lie close to the diagonal arise more likely due to noise fluctuations, whereas features far away from the diagonal are statistically more significant. The final chosen evaluation points for the persistent Betti functions can be seen in Table B.1. Over the course of the analysis, we tested several methods of choosing evaluation points, including just setting them by hand. While different sets of evaluation points sometimes slightly change the size of the contours, they are all internally consistent and yield very similar results. In particular, completely disregarding b_0 and $b_0^{\rm per}$ does not change the marginalised error on S_8, but helps in constraining the most remote regions of the $[\Omega_{\rm m}-\sigma_8]$ contours.

Appendix C: Evaluating the emulator accuracy

In this section, we further discuss the accuracy of the GPR emulator introduced in Sect. 2.4. As described therein, we evaluate the performance via a leave-one-out cross-validation test, and present our results in Fig. C.1 for the persistent Betti functions. As can be seen for the KV450-like case (upper panel), the emulator manages to predict most function values with a 5% accuracy. While some points are less precise, with inaccuracies ranging up to 15%, the mean inaccuracy (represented by the black line) stays well below 1%, meaning that the emulator neither systematically under- nor over-predicts any point of the Betti functions. We next assess the impact of the emulator uncertainty on the inferred cosmological parameters with an MCMC analysis in which we set the emulator covariance $C_{\rm e}$ to zero. As reported in Fig. C.2, this test confirms that the emulator covariance plays a non-negligible role at the moment since in that case the constraining power is increased for all parameters, and notably the constraints on S_8 become tighter by ∼5%.

We applied the same procedure to the *Euclid*-like simulations presented in Sect. 4 and show the cross-validation test in the bottom panel of Fig. C.1. Here, the emulator inaccuracies are significantly higher: While the mean fractional error in the cross-validation test still stays within a few percent, individual deviations often surpass 15%, some points even exceeding 50% fractional error. When compared with the fractional statistical error (the grey band in the figure), we observe that it is also larger compared to the KV450-like analysis. The resulting effect on the cosmological parameter analysis can be seen in Fig. C.3. While the peak statistics are only slightly affected in that case, the difference is significant for persistent Betti functions. In particular, the constraints on S_8, $\Omega_{\rm m}$, and w_0 improve by 4, 9, and 15%, respectively.

The fact that a less noisy measurement yields a larger fractional statistical error might be counter-intuitive, but it can be understood in the following way: In contrast to other measurements, including two-point correlation functions, the presence of shape noise does not only lead to larger measurement errors, but it also changes the expectation value of the measurement itself. Due to the presence of noise, an additional number of features (or peaks) is added to each measurement. This number of pure noise peaks is, to leading order, independent of cosmology, and since adding a constant number to a series of measurements does not affect its variance, the total sampling variance σ is independent of the amount of noise features. When the shape noise is lowered in our Stage IV survey simulations, the variance stays approximately constant, but the amplitude of the data vector decreases, causing the fractional statistical error to increase, as seen by comparing the size of the grey bands in both panels of Fig. C.1. We carried out an extra test in which the intrinsic shape noise of every galaxy in the KV450-like setup was reduced by 90%, and saw the same trend: With lower noise levels, the fractional statistical error increased almost by a factor of two in that case.

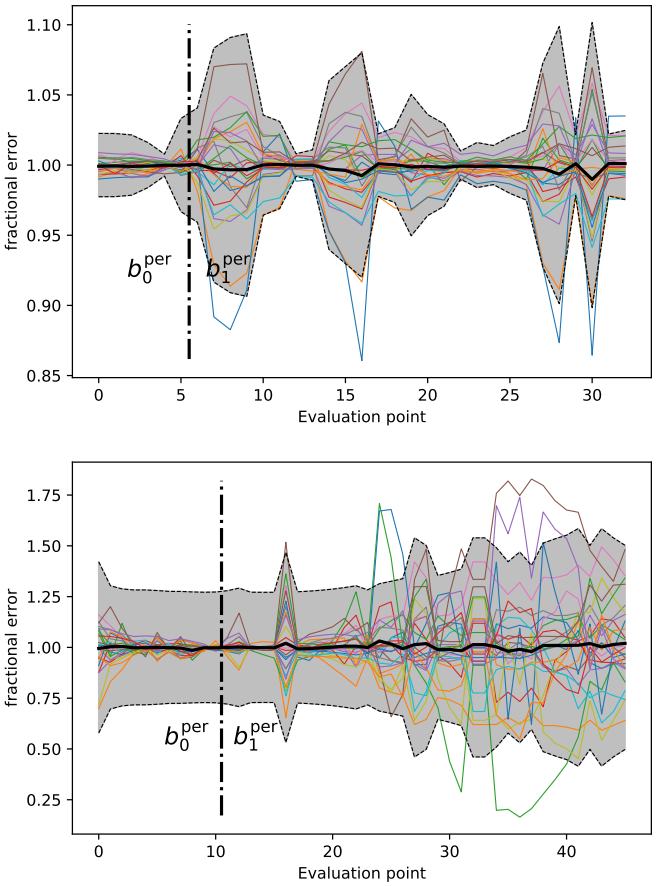


Fig. C.1. Relative accuracy of the emulator for predicting the persistent Betti functions of a KV450-like (*top*) and a *Euclid*-like (*bottom*) survey. The thin-coloured lines represent the ratio between predicted and measured values for the 26 different cosmologies, estimated from a cross-validation test. The x-axis lists the points at which the functions are evaluated (see Table B.1 for their numerical values). The thick black line corresponds to the mean over all lines. The grey shaded area corresponds to the 1σ standard deviation of the covariance matrix extracted from SLICS.

This also explains why the emulator has more difficulties in maintaining a high level of accuracy: The constant noise contribution dilutes the relative variations due to cosmology, which makes the interpolation easier for the GPR. In the *Euclid*-like mocks, the emulator therefore accounts for a large fraction of the error budget. These emulator inaccuracies, and how to improve upon them, will be subject to further investigation.

While we are not yet able to reduce modelling uncertainties with the cosmo-SLICS, this will certainly be possible in the future. Figure A.2 of H+19 shows that for two-point correlation functions, the uncertainty of the emulator falls to the order of a percent when training it on a set of functions that has been

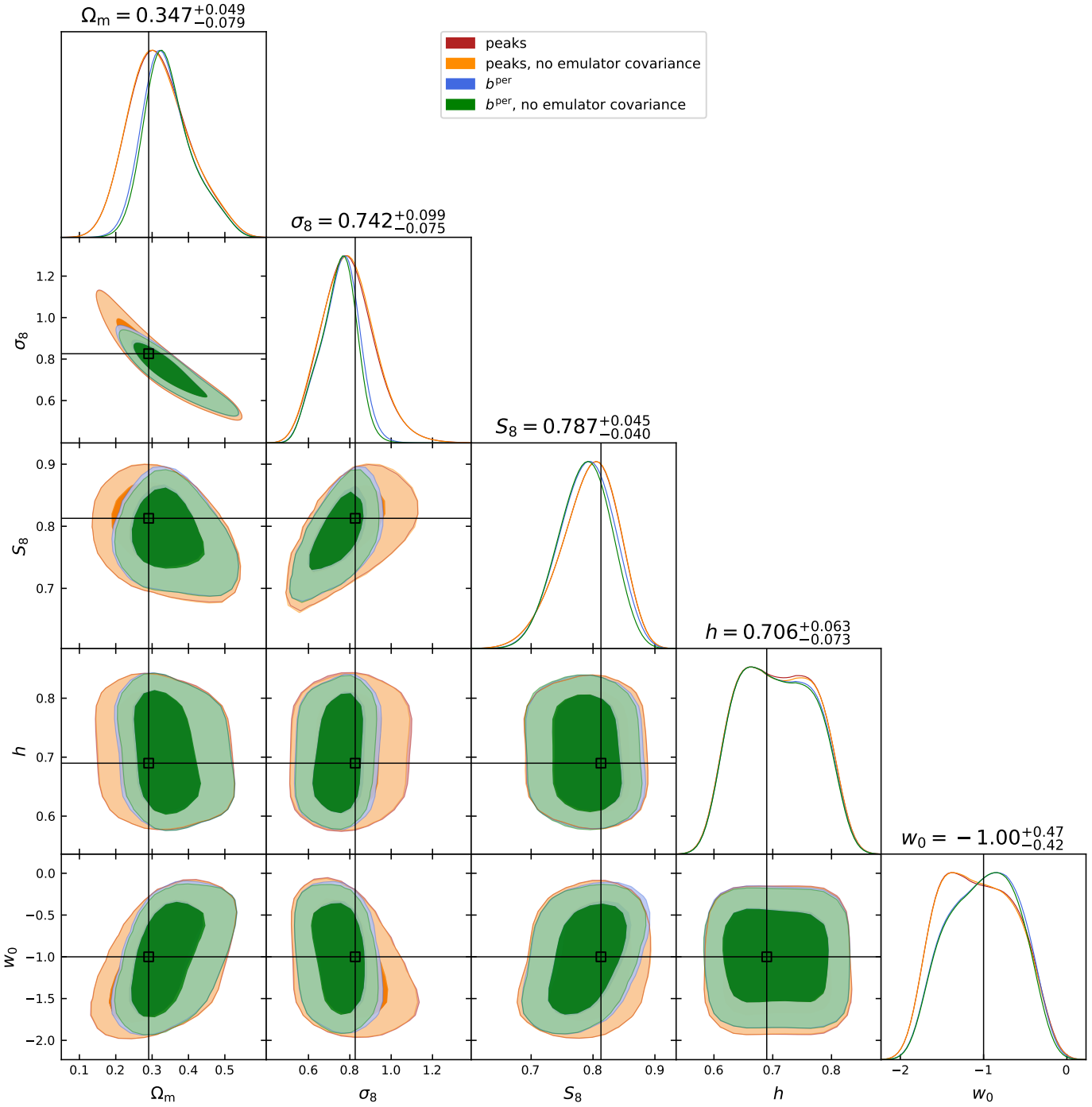


Fig. C.2. Parameter constraints from the persistent Betti functions and the peak statistics as presented in Fig. 5 (blue and red, respectively), here compared to the case where the covariance of the emulator C_e is set to zero in the MCMC analyses (green and orange). We note that for the peaks, the emulator covariance plays a negligible role.

modelled for 250 separate cosmologies. With larger projects like BACCO (Angulo et al. 2020), where simulations are conducted for 800 different cosmologies, this accuracy is definitely achievable if a ray-tracing is performed. In that case, it is interesting to investigate the impact of a better emulator on our analysis. This can be achieved simply by performing again a cosmological parameter analysis where we set $C_e = 0$, as shown in Figs. C.2 and C.3.

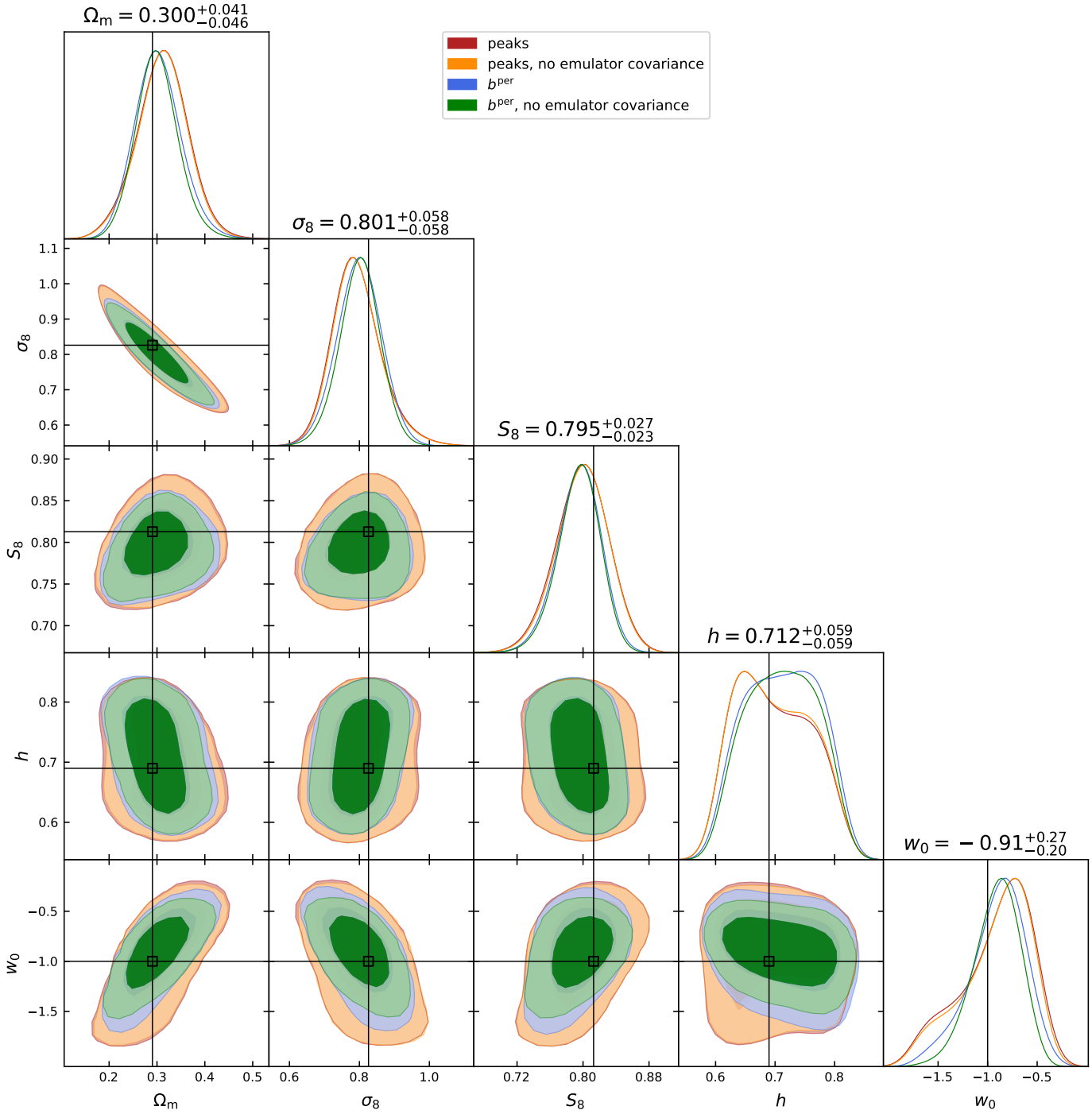


Fig. C.3. Same as Fig. C.2, here in the *Euclid*-like setup. We note that for the peaks, the emulator covariance plays a sub-dominant role. The results also indicate that, given a better emulator, persistent Betti numbers will be able to constrain the equation of state of dark energy without the need for tomographic analyses.

Appendix D: Treatment of biases by noise terms

To test the sensitivity to number density and shape noise, we created different galaxy catalogues of SLICS, where we both increased and reduced the shape noise of galaxies by 10%, and one where we reduced the number density by 10% by removing galaxies at random. We used these modified catalogues to extract the mean of the persistent Betti functions $b^{\rm per}$ as well as their covariance matrix. We then performed a cosmological parameter inference as in Sect. 3, where in each case the $\boldsymbol{x}_{\rm predicted}$ were provided by the emulator that has been trained on the unaltered cosmo-SLICS. As can be seen in Fig. D.1, a 10% change in these parameters roughly corresponds to a $\sim 1\sigma$ shift in the posterior distribution of S_8. Additionally, the posterior distributions for both $\Omega_{\rm m}$ and σ_8 are shifted by $\sim 3\sigma$ for the case of the lowered number density. It is therefore critical to reproduce these properties in the simulations in order to avoid very significant biases in the cosmological inference. In an actual cosmological data analysis, one should use mock data with a similar design to those used in this analysis, namely where galaxies have been distributed on the footprint exactly according to their positions in the data such as to reproduce number density, and additionally keeping their associate ellipticity.

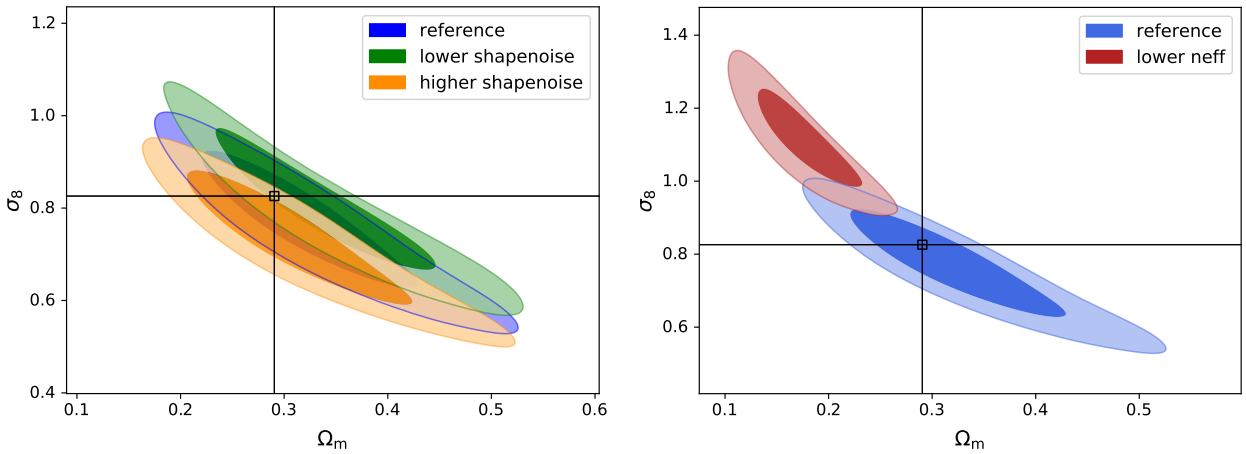


Fig. D.1. Effect on the inferred cosmological parameter in a KV450-like survey from varying shape noise by ±10% (*left*) and lowering the number density by 10% (*right*). The solid lines correspond to the fiducial values of SLICS.

C.2. Persistent homology in cosmic shear: A tomographic analysis of DES-Y1

Astronomy & Astrophysics manuscript no. main
August 19, 2022

Persistent homology in cosmic shear II: A tomographic analysis of DES-Y1

Sven Heydenreich[1], Benjamin Brück[2], Pierre Burger[1], Joachim Harnois-Déraps[3,4], Sandra Unruh[1,5], Tiago Castro[6,7,8], Klaus Dolag[9,10], Nicolas Martinet[11]

[1] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[2] ETH Zürich, Department of Mathematical Sciences, Rämistrasse 101, 8092 Zürich, Switzerland
[3] School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, NE1 7RU, Newcastle-upon-Tyne, UK
[4] Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF
[5] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
[6] Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131, Trieste, Italy
[7] Institute for Fundamental Physics of the Universe, via Beirut 2, 34151, Trieste, Italy
[8] INFN – Sezione di Trieste, I-34100 Trieste, Italy
[9] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstr.1, 81679 München, Germany
[10] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Straße 1, 85741 Garching, Germany
[11] Aix-Marseille Univ, CNRS, CNES, LAM, Marseille, France
e-mail: sven@astro.uni-bonn.de

Version August 19, 2022; received xxx, accepted yyy

ABSTRACT

We demonstrate how to use persistent homology for cosmological parameter inference in a tomographic cosmic shear survey. We obtain the first cosmological parameter constraints from persistent homology by applying our method to the first-year data of the Dark Energy Survey.
To obtain these constraints, we analyse the topological structure of the matter distribution by extracting persistence diagrams from signal-to-noise maps of aperture masses. This presents a natural extension to the widely used peak count statistics. Extracting the persistence diagrams from the cosmo-SLICS, a suite of N-body simulations with variable cosmological parameters, we interpolate the signal using Gaussian processes and marginalise over the most relevant systematic effects, including intrinsic alignments and baryonic effects.
For the structure growth parameter, we find $S_8 = 0.747^{+0.025}_{-0.031}$, which is in full agreement with other late-time probes. We also constrain the intrinsic alignment parameter to $A = 1.54 \pm 0.52$, which constitutes a detection of the intrinsic alignment effect at almost 3σ.

Key words. gravitational lensing – weak, cosmology – cosmological parameters, methods – topological data analysis

1. Introduction

In the past decades, weak gravitational lensing has emerged as an indispensable tool for studying the large-scale structure (LSS) of the Universe. Weak lensing primarily relies on accurate shape and distance measurements of galaxies. Ongoing and recently completed surveys have provided the community with a sizeable amount of high-quality data; for example, the Kilo Degree Survey (KiDS, de Jong et al. 2013), the Dark Energy Survey (DES, Flaugher 2005), and the Hyper Suprime-Cam Subaru Strategic Program (HSC, Aihara et al. 2018). Further surveys are scheduled to start observing in the next years; they will probe deeper and larger areas enabling measurements of cosmological parameters with sub-per cent accuracy; for example, the Vera Rubin Observatory's Legacy Survey of Space and Time (LSST, Ivezic et al. 2008), the Euclid survey (Laureijs et al. 2011), and the Nancy Grace Roman Space Telescope (RST) survey (Spergel et al. 2013). These upcoming surveys are of special relevance for solving tensions related to measurements of the structure growth parameter $S_8 = \sigma_8\sqrt{\Omega_\mathrm{m}/0.3}$ (Hildebrandt et al. 2017; Planck Collaboration et al. 2020; Joudaki et al. 2020; Heymans et al. 2021; Abbott et al. 2022), which is defined along the main degeneracy direction in conventional weak lensing studies. Here, Ω_m is the dimensionless matter density parameter and σ_8 parametrises the amplitude of the matter power spectrum. Improved data and independent analysis choices are crucial to determine whether this tension is due to new physics, a statistical fluctuation, or the manifestation of unknown systematics. For example, Joudaki et al. (2017) showed that the current tension in S_8 between the CMB and the local Universe could be lifted when allowing for a dynamical dark energy model, meaning that measuring the equation of state of dark energy is of the utmost importance in the next decades.

Shear two-point statistics have emerged as the prime analysis choice for cosmic shear as they present a num-

Article published by EDP Sciences, to be cited as https://doi.org/10.1051/0004-6361/202243868

ber of key advantages (e.g. Secco et al. 2022; Hikage et al. 2019; Asgari et al. 2021). Such statistics are physically motivated by the fact that they describe the early Universe almost perfectly. The late Universe, however, contains a considerable amount of non-Gaussian information that is not captured by two-point statistics, such that jointly investigating second- and higher-order statistics increases the constraining power on cosmological parameters (see, e.g. Bergé et al. 2010; Pyne & Joachimi 2021). This additional information is currently explored with a variety of analysis tools, which either use analytical models (e.g. Halder et al. 2021; Gatti et al. 2021; Burger et al. 2022) or rely on large suites of numerical simulations. For this work, the most relevant examples for a simulation-based analysis are peak statistics (Martinet et al. 2021b, and references therein, hereafter M+21) and Minkowski functionals (e.g. Shirasaki & Yoshida 2014; Petri et al. 2015; Parroni et al. 2020), which are both based on aperture mass maps constructed from shear fields. M+21 also showed that a joint analysis of peaks and two-point correlation functions (2PCF) improves cosmological constraints on S_8, Ω_m, and the dark energy equation-of-state parameter w_0 by 46%, 57%, and 68%, respectively. Zürcher et al. (2021) showed that a joint analysis using 2PCF with Minkowski functionals, a topological summary statistic, on aperture mass maps increases the figure of merit in the Ω_m–σ_8 plane by a factor of 2. We note that there exist many other promising simulation-based methods such as bayesian hierarchical forward modelling (Porqueres et al. 2021, 2022), likelihood-free inference (Jeffrey et al. 2021), or the scattering transform (Cheng et al. 2020).

In this paper, we focus on persistent homology, a topological method that combines the advantages of peak statistics and Minkowski functionals but also captures information about the environment of topological features. Persistent homology specialises in recognising persistent topological structures in data and we refer the interested reader to a recent review written by Wasserman (2018), who highlights its diverse applications in various fields. Following early concepts about persistent homology and Betti numbers in cosmology (van de Weygaert et al. 2013), several groups have formalised the approach (Sousbie 2011; Pranav et al. 2017; Feldbrugge et al. 2019; Pranav 2021). In particular, Kimura & Imai (2017) were the first to show that the hierarchical topological structure of the galaxy distribution decreases with increasing redshift using small patches of Sloan Digital Sky Survey (SDSS). More recently, Xu et al. (2019) developed an effective cosmic void finder based on persistent homology, while Kono et al. (2020) detected baryonic acoustic oscillations in the quasar sample from the extended Baryon Oscillation Spectroscopic Survey in SDSS. Moreover, Biagetti et al. (2020, 2022) showed with simulations that persistent homology is able to identify primordial non-Gaussian features. Heydenreich et al. (2021, hereafter H+21) performed a mock analysis using persistent homology on cosmic shear simulations, highlighting its potential to break the degeneracy between S_8 and w_0.

Persistent homology summarises the topological structure of data in so-called persistence diagrams. There are different methods for performing statistical analyses on such diagrams (see Sect. 3.2). In H+21, we worked with persistent Betti numbers. In this work, we opted for 'heatmaps', which constitute a more robust statistic for persistence diagrams. We extract heatmaps from a series of mock data that match the DES-Y1 survey properties (Flaugher 2005; Harnois-Déraps et al. 2021, hereafter HD+21), including a Cosmology Training Set, a Covariance Training Set, and a suite of Systematics Training Set, constructed from the SLICS (Harnois-Déraps & van Waerbeke 2015), the cosmo-SLICS simulations (Harnois-Déraps et al. 2019) and the Magneticum hydrodynamical simulations (Biffi et al. 2013; Saro et al. 2014; Steinborn et al. 2015, 2016; Dolag 2015; Teklu et al. 2015; Bocquet et al. 2016; Remus et al. 2017; Castro et al. 2018, 2021). Following HD+21, we then train a Gaussian process regression (GPR) emulator, which is fed to a Markov Chain Monte Carlo (MCMC) sampler to obtain cosmological parameter estimates. We significantly expand on the results from H+21 by including the main systematic effects related to cosmic shear analyses, namely photometric redshift uncertainty, shear calibration, intrinsic alignment of galaxies, baryon feedback and masking. These systematics, particularly baryon feedback and intrinsic alignments, account for 25% of our reported final error budget. Furthermore, as introduced in M+21, our results are obtained for a tomographic topological data analysis where we include the cross-redshift bins analyses. This leads to the first cosmological parameter constraints obtained from persistent homology based on analysing cosmic shear data, here provided by the DES year-1 survey (Abbott et al. 2018).

The paper is organised as follows: In Sect. 2 we describe the data and simulations; the theoretical background on persistent homology, a description of our data compression methods, the formalism for the two-point statistics and the cosmological parameter estimation are presented in Sect. 3. In Sect 4 we discuss our mitigation strategies for systematic effects and show the validation of our pipeline in Sect. 5. We finalise our work with the results shown in Sect. 6 and our discussion in Sect. 7.

2. Data and simulations

2.1. DES-Y1 data

We use in this work the public[1] Year-1 data released by DES presented in Abbott et al. (2018, DES-Y1 hereafter). The primary weak lensing data consist of a galaxy catalogue in which positions and ellipticities are recorded for tens of millions of objects, based on observations from DECam mounted at the Blanco telescope at the Cerro Tololo Inter-American Observatory (Flaugher et al. 2015). The galaxies selected in this work match those of Troxel et al. (2018, hereafter T+18) and HD+21, applying the FLAGS SELECT, METACAL, and the REDMAGIC filters to the public catalogues, yielding a total unmasked area of 1321 deg^2 and 26 million galaxies.

The shear signal $\gamma_{1/2}$ is inferred from the METACALIBRATION technique (Sheldon & Huff 2017), which further provides each galaxy with a METACAL response function S_i that must be included in the measurement. As explained in T+18, this method requires a prior on an overall multiplicative shear correction of $m \pm \sigma_m = 0.012 \pm 0.023$, which we then use to calibrate the measured galaxy ellipticities as $\epsilon_{1/2} \rightarrow \epsilon_{1/2}(1+m)$. We then assume that these ellipticities are an unbiased estimator for the shear γ.

Following T+18, the galaxy sample is further split into four tomographic bins based on their individual estimated

[1] DES-Y1 catalogues: des.ncsa.illinois.edu/releases/dr1

photometric redshift Z_B, which is measured with the BPZ method (Benítez 2000). At this point, the redshift distribution of the four tomographic populations are estimated with the 'DIR'[2] method, following Joudaki et al. (2020) and HD+21. As argued in these two references, the DIR approach is more robust to potential residual selection effects in their training sample than the DES-Y1 BPZ stacking method presented in Hoyle et al. (2018). Although it could also be affected by incomplete spectroscopy and colour preselection (Gruen & Brimioulle 2017), bootstrap resampling of the spectroscopic samples points towards a significantly smaller uncertainty in the mean redshift of the populations, achieving $\sigma_z = 0.008, 0.014, 0.011$ and 0.009 for tomographic bins 1..4, respectively (Joudaki et al. 2020). Despite these important differences, the DIR $n(z)$ is consistent with the fiducial estimate presented by the DES collaboration. It brings in excellent agreement the DES-Y1 and the KV-450 cosmic shear data (Hildebrandt et al. 2020, also based on the DIR method). Asgari et al. (2020) showed that the inferred S_8 value is affected by less than 1σ, which is certainly a considerable effect but causes no internal tension between the two methods of redshift estimation.

2.2. Mock galaxy catalogues

The analysis presented in this work largely follows the simulation-based inference methods of HD+21, which completely relies on numerical weak lensing simulations for the cosmology inference, the estimation of the uncertainty and the mitigation of systematics and secondary effects. Most of the mock data used in this work has been presented in HD+21, which we review in this section.

2.2.1. Cosmology Training Set

This set of simulations is used to model the dependence of the signal on cosmology. Based on the cosmo-SLICS (Harnois-Déraps et al. 2019), it consists of weak lensing light cones sampling 26 points in a wCDM cosmological model (i.e. cold dark matter with dark energy beyond the cosmological constant Λ), where 25 points are distributed on a Latin Hypercube, covering the ranges $\Omega_\mathrm{m} \in [0.10, 0.55]$, $S_8 \in [0.60, 0.90]$, $h \in [0.6, 0.82]$ and $w_0 \in [-2.0, -0.5]$, where h is the reduced Hubble parameter. The last point is set manually to a fiducial, ΛCDM cosmology. Each node consists of two independent N-body simulations produced by CUBEP^{3}M (Harnois-Déraps et al. 2013), with initial conditions designed such as to suppress the sampling variance. The code follows the non-linear evolution of 1536^3 particles in a 505 h^{-1} Mpc box, producing between 15 and 28 mass sheets of co-moving thickness equivalent to half the box size, filling up a $10 \times 10\,\mathrm{deg}^2$ light cone to $z = 3.0$. Random orientations and shifting are introduced in this process such that a total of 25 pseudo-independent light cones are generated per N-body run. Five of these are used in the current paper, out of 25, which is sufficient to model the statistics within the DES-Y1 precision (as in HD+21). Validation tests revealed that the third line of sight from the first N-body seed is a statistical outlier: for example, the standard deviation of the convergence σ_κ differs from the mean of the full cosmo-SLICS light cones by more than 4σ. Due to the limited size of our training sample, this particular line of sight could bias our cosmological model. We thus skipped over it and verified afterwards that the results are not strongly affected by this choice, although it slightly improves the accuracy on S_8 during the validation test. In total, the Cosmology Training Set consists of $26 \times 9 = 234$ survey realisations.

2.2.2. Covariance Training Set

This suite is mainly used to estimate the sampling covariance in the data vector. Based on the SLICS (Harnois-Déraps & van Waerbeke 2015), it is produced from 124 fully independent N-body realisations, with the same mass resolution and simulation volume as the cosmo-SLICS. All carried out at the same cosmology, these light cones started from different noise realisations of the initial conditions, thereby sampling the statistical variance in the data. The mean over all measurements from the Covariance Training Set is also independent of the Cosmology Training Set and well converged towards the ensemble average, making this an ideal data set with which we validate our cosmology inference pipeline later on.

2.2.3. Systematics Training Set - Mass resolution

The force resolution of N-body simulations is limited by the number of particles, the choice of softening length and the force accuracy setting. This inevitably translates into a decrease in the clustering of dark matter in the highly non-linear scales, which in turn affects the statistics under study. The SLICS-HR are a suite of high-resolution simulations introduced in Harnois-Déraps & van Waerbeke (2015), in which the force accuracy of CUBEP^{3}M has been significantly increased, yielding 5 light cones with more accurate mass densities. As detailed in Sect. 4, we verify that our training data are not strongly affected by this known limitation.

2.2.4. Systematics Training Set - Baryons

Baryonic feedback processes from sustained stellar winds, supernovae and active galactic nuclei are known to redistribute the matter around over-dense regions of the Universe in a manner that directly affects the weak lensing measurements (Semboloni et al. 2011). If left unmodelled, these processes will significantly bias the inferred cosmology in analyses based on 2PCF or non-Gaussian statistics (e.g. Coulton et al. 2020; Zürcher et al. 2021; Martinet et al. 2021a). In this work, our approach consists of measuring our statistics in hydrodynamical simulations in which the baryon feedback can be turned on and off. The relative impact on the data vector is then used to model the effect of baryons on our statistics. As in HD+21, we use the Magneticum simulations[3] to achieve this, more precisely the Magneticum Run-2 and Run-2b (Castro et al. 2021), in which stellar formation, radiative cooling, supernovae and AGN feedback are implemented in cosmological volumes of 352 and 640 h^{-1} Mpc, respectively, with a spatial resolution that is high enough to capture the baryonic effects at scales relevant to our study. The adopted cosmology is consistent with the SLICS cosmology, with $\Omega_\mathrm{m} = 0.272$, h

[2] This method relies on the direct calibration of the $n(z)$ from a sub-sample of DES-Y1 galaxies for which external spectroscopic data are available. See Lima et al. (2008) for more details.

[3] www.magneticum.org

$= 0.704$, $\Omega_\mathrm{b} = 0.0451$, $n_\mathrm{s} = 0.963$, and $\sigma_8 = 0.809$. These simulations reproduce a number of key observations, including many statistical properties of the large-scale, intergalactic, and intercluster medium (see Hirschmann et al. 2014; Teklu et al. 2015; Castro et al. 2018, for more details). Moreover, the resulting overall feedback is consistent with that of the BAHAMAS simulations (McCarthy et al. 2017), which are based on a completely independent sub-grid calibration method. The Baryons Training Set and their dark-matter-only counterpart are used to inspect the impact of baryonic physics on the data vector, from which we extract a correction factor used to forward-model the effect on dark-matter-only simulations. Full details on the treatment of the systematics are presented in Section 4.

2.2.5. Systematics Training Set - Photometric redshifts

The redshift distribution of the data is known to a high precision within the DIR method however, the residual uncertainty must be accounted for in the analysis. For this, we use the mocks described in HD+21 in which the $n(z)$ has been shifted by a small amount in order to study the impact on the signal. These sample at ten points the posterior of the expected shifts in the mean redshifts of the DIR method itself (Joudaki et al. 2020), and in each case, we construct 10 full survey realisations at the cosmo-SLICS fiducial cosmology, from which we extract our statistics. This approach allows us to measure the derivative of the persistent homology statistics with respect to shifts in $\mathrm{d}z$. The priors on $\mathrm{d}z$ are listed in Table 1.

2.2.6. Systematics Training Set - Intrinsic Alignments

The assumption that the observed shapes of galaxies are randomly aligned in absence of foreground lensing matter fails to account for their intrinsic alignment (IA), an important contribution that arises from a coupling between their shapes and the large-scale structure they are part of (for a review see Joachimi et al. 2015). This important secondary signal tends to counteract the cosmic shear signal, which can therefore interfere with the cosmological inference. Although there exist analytical models to describe this effect for two-point functions, higher-order statistics must rely on IA-infused simulations to account for this important effect. In this work, we use the infusion method presented in Harnois-Déraps et al. (2022), where intrinsic galaxy shapes are linearly coupled with the projected tidal field, consistent with the non-linear alignment model of Bridle & King (2007). Although the redshift distribution of these mocks exactly follows that of the data, their construction requires that the galaxy ellipticities linearly trace the simulation density fields, whose positions therefore no longer replicate that of the DES-Y1 data. These mocks have no masking, nor METACALIBRATION responses included and are therefore used to estimate the relative impact of IA on our persistent homology measurements. They have an IA amplitude that is allowed to vary, as controlled by the A_IA parameter. We measure the persistent homology statistics from 50 cosmo-SLICS light cones at the fiducial cosmology, for values of $A_\mathrm{IA} \in [-5.0, 5.0]$, and use these to construct a derivative, similar to the way we handle the photometric redshift uncertainty.

2.2.7. Creating the galaxy catalogues

The output of each simulation is a series of 100 deg^2 lensing planes that serve to assign convergence (κ) and shear ($\gamma_{1/2}$) to copies of the DES-Y1 data. As described in HD+21, the survey footprint is segmented into 19 regions, or tiles, which all fit inside our simulated maps. The summary statistics are computed individually on each tile and combined afterwards to construct the data vector. In this construct, the galaxy positions, ellipticities $\epsilon_{1/2}$ and METACALIBRATION weights S_i in the mock data exactly match that of the real data, avoiding possible biases arising in non-Gaussian statistics when these differ (see e.g. Kacprzak et al. 2016, Appendix D of H+21). Mock ellipticities are obtained by rotating the observed ellipticities by a random angle and combining the resulting randomised signal $\boldsymbol{\epsilon}_\mathrm{n}$ with the simulated (noise-free) reduced shear $\boldsymbol{g}$ via:

$$\boldsymbol{\epsilon} = \frac{\boldsymbol{\epsilon}_\mathrm{n} + \boldsymbol{g}}{1 + \boldsymbol{\epsilon}_\mathrm{n}\boldsymbol{g}^*} \,, \tag{1}$$

where bold symbols denote complex numbers (for example, $\boldsymbol{g} = g_1 + \mathrm{i}g_2$). We calculate the reduced shear as $\boldsymbol{g} = \boldsymbol{\gamma}/(1-\kappa)$. In total, we compute 10 shape-noise realisations for every simulated survey realisation, each using a different random seed in the rotation. This procedure allows us to average out a large part of the fluctuations introduced by the shape noise, improving both our predictions and our estimate of the sample covariance while preserving the data noise levels. Redshifts are assigned to every simulated galaxy by sampling from the DIR redshift distribution corresponding to the tomographic bin they belong to.

2.3. Calculating maps of aperture masses

As in H+21, we perform our computations on signal-to-noise maps of aperture masses (Schneider 1996; Bartelmann & Schneider 2001), computing the signal $\mathcal{M}_\mathrm{ap}(\boldsymbol{\theta})$ and noise $\sigma\left(\mathcal{M}_\mathrm{ap}(\boldsymbol{\theta})\right)$ on a grid as:

$$\mathcal{M}_\mathrm{ap}(\boldsymbol{\theta}) = \frac{1}{n_\mathrm{gal}\sum_i w_i S_i} \sum_i Q(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|) w_i \epsilon_\mathrm{t}(\boldsymbol{\theta}_i; \boldsymbol{\theta}) \,, \tag{2}$$

$$\sigma\left(\mathcal{M}_\mathrm{ap}(\boldsymbol{\theta})\right) = \frac{1}{\sqrt{2} n_\mathrm{gal}\sum_i w_i S_i} \sqrt{\sum_i |w_i \epsilon(\boldsymbol{\theta}_i)|^2 Q^2(|\boldsymbol{\theta}_i - \boldsymbol{\theta}|)} \,, \tag{3}$$

where the w_i are optional weights assigned to measured galaxy ellipticities (set to 1.0 in this work), S_i are the respective responses calculated by the METACALIBRATION shear estimator (T+18), and the tangential component of the shear $\epsilon_\mathrm{t}(\boldsymbol{\theta}_i; \boldsymbol{\theta})$ is calculated via

$$\epsilon_\mathrm{t}(\boldsymbol{\theta}_i; \boldsymbol{\theta}) = -(\epsilon_1 + \mathrm{i}\epsilon_2) \frac{(\boldsymbol{\theta}_i - \boldsymbol{\theta})^*}{(\boldsymbol{\theta}_i - \boldsymbol{\theta})} \,. \tag{4}$$

We then compute the signal-to-noise map (S/N map) of aperture masses as the ratio between the two quantities. As before, we use the following Q-filter function (Schirmer et al. 2007; Martinet et al. 2018, hereafter M+18):

$$Q(\theta) = \left[1 + \exp\left(6 - 150\frac{\theta}{\theta_\mathrm{ap}}\right) + \exp\left(-47 + 50\frac{\theta}{\theta_\mathrm{ap}}\right)\right]^{-1} \times \left(\frac{\theta}{x_\mathrm{c}\theta_\mathrm{ap}}\right)^{-1} \tanh\left(\frac{\theta}{x_\mathrm{c}\theta_\mathrm{ap}}\right) \,, \tag{5}$$

with a concentration index of $x_c = 0.15$ (Hetterscheidt et al. 2005), which was chosen to optimally select the mass profiles of dark matter halos (Navarro et al. 1997). For the filter radius we choose $\theta_{\rm ap} = 12.5'$. As in H+21, we compute the S/N maps by distributing both galaxy ellipticities ϵ_i and their squared moduli $|\epsilon_i|^2$ for each tile on a (600×600) pixel grid, and perform the convolutions in Eqs. (2) and (3) via a Fast Fourier-Transform. Contrary to previous work, we use a cloud-in-cell algorithm to distribute the galaxy ellipticities on a grid, yielding more accurate results for small scales when dealing with high-quality data.

As shown in M+21, the traditional approach of computing the aperture mass statistics for individual tomographic bins only (hereafter auto-bins) does not yield optimal results. Instead, we perform the computation for all combinations of tomographic bins by concatenating the respective galaxy catalogues (cross-bins). This approach allows us to extract additional information about correlated structures along the line of sight. For example, a massive, nearby galaxy cluster can be detected as a peak in the S/N maps for tomographic bins 1 and 2. However, if we were only to analyse persistence heatmaps of the two respective bins, both would register the cluster as a peak, but the information that the peak is at the same position in both maps would be lost. To utilise this information, we also need to analyse the S/N map of a combination of both bins. Based on the four fiducial DES-Y1 redshift bins, this optimised method leads to 15 bin combinations (1, 2, 3, 4, 1∪2, ..., 3∪4, ..., 1∪2∪3∪4) from which we extract heatmaps.

As the galaxies in our mock data follow the exact positions of the real galaxy catalogue, they are subject to the overall survey footprint and internal masked regions. We only want to consider the parts of our S/N maps where we have sufficient information from surrounding galaxies, therefore we construct our own mask in the following way: We combine the galaxy catalogues of all four tomographic bins and distribute these galaxies on a grid. Then we mask all pixels of the tile where the effective area containing galaxies within the aperture radius $\theta_{\rm ap}$ is less than 50%. In particular, we mask the boundary of each tile to ensure that neighbouring tiles are treated as independent in the persistence calculations (compare H+21). This mask is then applied to every combination of tomographic bins of the respective survey tile.

With this method, we compute the S/N maps for each of the 19 tiles of the DES-Y1 survey footprint, for each of the 15 tomographic bins and for each of the 10 shape noise realisations. The next section describes how cosmological information is extracted from these maps with statistics based on persistent homology.

3. Methods

We use methods from persistent homology to quantify the statistical properties of S/N maps of aperture masses and analyse their dependence on the underlying cosmological parameters. The main idea can be described as follows:

We take a S/N map of aperture masses and apply a threshold to that map. We then cut off all parts where the value of the S/N map exceeds that threshold (compare Fig. 1). This gives rise to two types of topological features. The first types are connected components, that is regions of low S/N that are surrounded by a region of higher S/N, which is above the cut-off threshold. These connected components correspond to local minima in the S/N map, which in turn correspond to an under-density in the matter distribution. The second type of topological features are holes, that is regions of high S/N that are above the cut-off threshold, with an environment of S/N that surrounds them and is lower than the cut-off threshold. These holes correspond to the local maxima of the S/N map, which indicate an overdensity in the underlying matter distribution.

When the cut-off threshold is gradually increased, these features change. Connected components start to show up (are born) once the threshold is higher than their minimum S/N value. At some higher threshold, the connected component will merge with a different connected component (or die)[4]. Similarly, an overdensity starts to form a hole once the cut-off threshold exceeds the S/N value of its environment. This hole is completely filled in once the threshold exceeds the maximal S/N value of the overdensity.

For each such topological feature, we write b for its birth (the threshold at which it is born) and d for its death (the threshold at which it dies). We plot the collection of all points (b, d) as a scatter plot, called the persistence diagram Dgm; we write Dgm_0 for the persistence diagram of connected components and Dgm_1 for the one of the holes (see Fig. 2). In particular, it is straightforward to recover the peak count statistics from this: The death of a hole corresponds to the maximal S/N value of an overdensity, so the set of deaths is the collection of all peaks in the S/N map. However, persistent homology offers one crucial advantage: The persistence of a feature, defined by $d - b$, yields information about how much a peak protrudes from its surrounding environment. In particular, features with a very small persistence are more likely to be caused by noise fluctuations, which can be taken into account in the following statistical analysis. Persistent homology offers a natural way to account for masked regions, which we describe in the next subsection. We denote the persistence diagrams that account for the presence of masked regions by Dgm_0^M and Dgm_1^M.

From the persistence diagrams Dgm_0^M and Dgm_1^M we then create so-called heatmaps by smoothing the diagrams with a Gaussian. Every point of the heatmap can now be used for statistical analysis of the persistent topological structure of the S/N maps of aperture mass. In the next two subsections, we give a slightly more formal introduction into these statistics derived from persistent homology and describe their application.

3.1. Persistent homology

In this section, we give a short overview of the aspects of persistent homology that we use in the present paper. More detailed explanations can be found in Sect. 2.3 of H+21. For a general introduction to the topic that is geared towards its applications in data science, see Chazal & Michel (2021) or Otter et al. (2017); further information about the mathematical background can be found in Oudot (2015).

Persistent homology is a technique from topological data analysis that allows summarising the topological features of a sequence of spaces. This is a versatile tool that can

[4] When two connected components merge, the one that was born at a lower threshold survives. This is known as the elder rule.

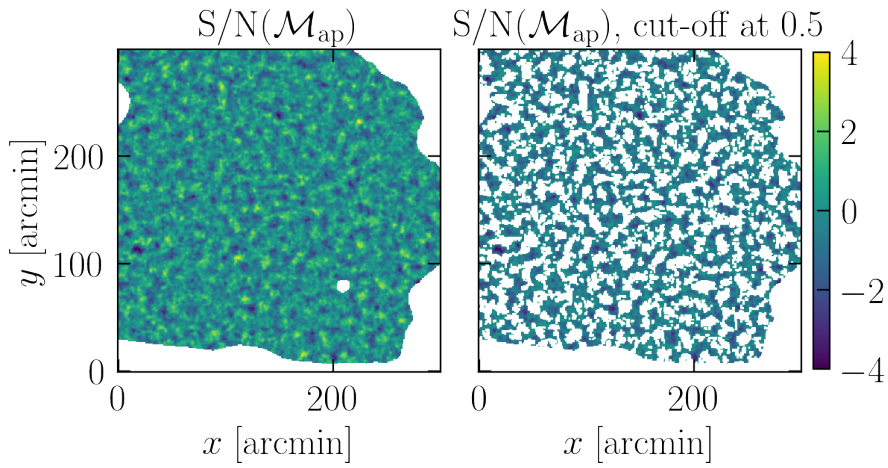


Fig. 1. Example signal-to-noise map of aperture masses for a $5 \times 5\,\mathrm{deg}^2$ sub-patch of one of the Covariance Training Set catalogue (left), and the same map when a threshold of 0.5 is applied (right). The white 'holes' in the right map correspond to local maxima of the map and give rise to the topological 'features' that are summarised in Dgm_1.

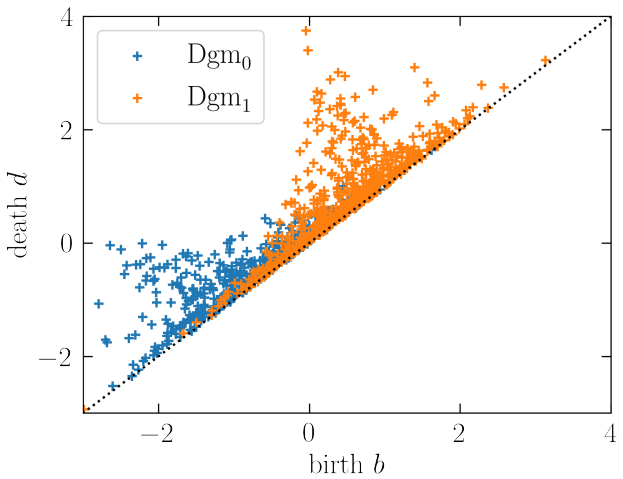


Fig. 2. Two persistence diagrams for the simulation shown in Fig.1. The blue crosses represent features of Dgm_0, the orange crosses represent features of Dgm_1. For visibility, only every 500-th feature is shown. We note that all points in this diagram lie above the diagonal.

be applied in many different settings. However, the application that is relevant for the present article is that persistent homology gives a summary of the topological features of a map $f : X \to \mathbb{R} \cup \{\pm\infty\}$, where X is in principle any (topological) space. (Here, the sequence of spaces is given by taking subsets of X that consist of points where the value of f lies below a certain threshold.) In our setting, X is a $(10 \times 10\,\mathrm{deg}^2)$ tile of the sky (which we interpret as a subset of $\mathbb{R}^2$ and represent by a (600×600) pixel grid) and f is the function that assigns to every point its S/N value as defined in Section 2.3. An example of this can be found in Fig. 1.

The persistent homology of each such map f can be summarised by two persistence diagrams $\mathrm{Dgm}_0 = \mathrm{Dgm}_0(f)$ and $\mathrm{Dgm}_1 = \mathrm{Dgm}_1(f)$. Each of these persistence diagrams is a collection of intervals $[b, d)$, where $b, d \in \mathbb{R} \cup \{\pm\infty\}$. As each such interval is determined by the two values $b < d$, one can equivalently see a persistence diagram as a collection of points (b, d) in $(\mathbb{R} \cup \{\pm\infty\})^2$ that lie above the diagonal. We call such a point (b, d) a feature of the persistence diagram; b is commonly called the birth and d is its death. Roughly speaking, the points in Dgm_0 correspond to the local minima of the function f, whereas the points in Dgm_1 correspond to the local maxima. In both cases, the difference $d - b$ of such a feature (b, d) is called its persistence and describes how much the corresponding extremum protrudes from its surroundings.

The actual computation of the Dgm_0 or Dgm_1 associated with an S/N map f is done as follows: As explained in Section 2.3, the S/N maps we compute are defined on a (600×600) pixel grid, and a subset of these pixels are masked. We set the value of every masked pixel to be $-\infty$ and compute the persistence diagrams Dgm_0 and Dgm_1 associated to this map[5] (see Fig. 2). As explained in H+21 (Sect. 2.3.3), relative homology offers a natural way to work with persistent homology in the presence of masks. The idea here is that a feature in Dgm_i that is of the form $(-\infty, d)$ corresponds to a minimum or maximum of f that originates from a masked area as these are the only points where f takes the value $-\infty$. This is why we do not actually work with Dgm_0 or Dgm_1 but instead define 'masked' persistence diagrams Dgm_0^M and Dgm_1^M as follows: The persistence diagram Dgm_0^M is obtained from Dgm_0 by simply removing all features coming from the masks, meaning all points of the form $(-\infty, d)$. To get Dgm_1^M from Dgm_1, we again start by removing all points of the form $(-\infty, d')$, but then for every feature of the form $(-\infty, d)$ in Dgm_0 (so those that got removed when transforming Dgm_0 into Dgm_0^M), we add a new feature (d, ∞) to Dgm_1^M[6].

3.2. Persistence Statistics

From the calculations described in the previous section, we obtain for each S/N map two persistence diagrams Dgm_0^M and Dgm_1^M. In order to carry out a statistical analysis of these persistence diagrams, one needs to be able to compute expected values and covariances. A priori, a persistence diagram is a particular collection of points in $(\mathbb{R} \cup \{\pm\infty\})^2$, and

[5] We use the `Cubical Complexes` module of the public software GUDHI (Dlotko 2020).

[6] Dgm_0^M and Dgm_1^M are the persistence diagrams associated to the persistence modules of the homology relative to the masked regions M. This is why Dgm_1^M is not simply obtained by removing all mask features from Dgm_1. For more explanations, see H+21, Sect. 2.3.3.

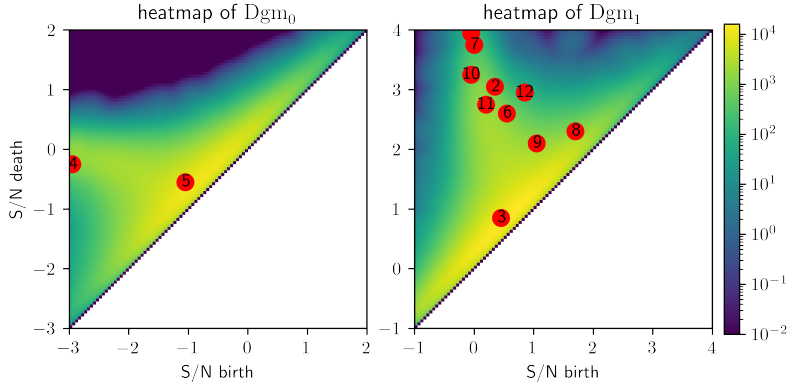


Fig. 3. Heatmap of the persistence diagram in Fig. 2 with a scaling parameter of $t = 0.2$ (for the computation of the Heatmaps, all features are taken into account, not just every 500-th as in Fig. 2). The red points correspond to the evaluation points that were chosen by the χ^2-maximiser outlined in Sect. 3.3. The extracted data vector can be seen in Fig. 4.

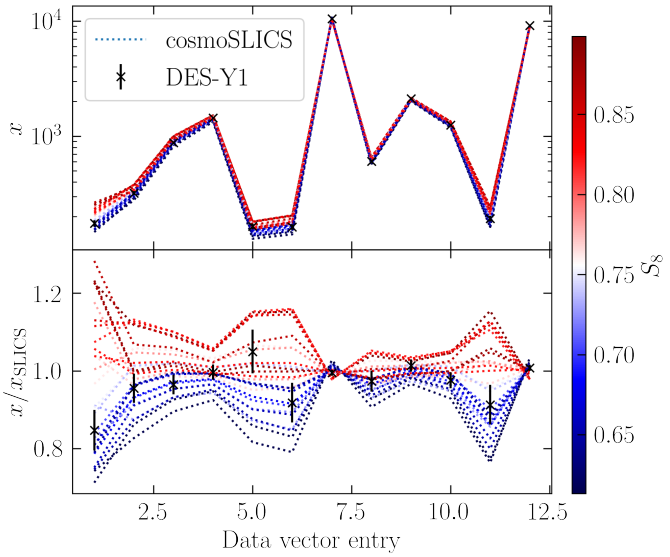


Fig. 4. Data vector for the individual cosmologies from our Cosmology Training Set, colour-coded by their respective value of S_8 (dotted lines) and the measured values in the DES-Y1 survey (black). For better visibility, the bottom panel shows the same data vector where all values are divided by the mean of our Covariance Training Set.

there is no canonical way of computing distances, sums and averages of such collections. There are different approaches to overcoming these difficulties. Most of them proceed by converting persistence diagrams into elements of a suitable vector space and then using tools for statistics and data analysis in this space. An overview of different options to perform statistics on persistence diagrams can be found in Chazal & Michel (2021), in particular Section 5.9 and in Pun et al. (2018). For this work, we tested three different approaches to the problem: persistent Betti numbers, persistence landscapes and heatmaps. All of these convert persistence diagrams into elements of certain function spaces.

Persistent Betti numbers are probably the most direct approach and were used in H+21. They represent a persistence diagram Dgm_i by the function $\beta_i : (\mathbb{R} \cup \{\pm\infty\})^2 \to \mathbb{R}$, where $\beta_i(x, y)$ is the number of points (b, d) in Dgm_i that lie to the 'upper left' of (x, y), meaning that $x \leq b$ and $d \leq y$. For more explanations about persistent Betti numbers, see H+21, Section 2.2.3 and Appendix B.

Persistent landscapes are a more elaborate alternative, introduced in Bubenik (2015) and already successfully used in applications, for example in Chittajallu et al. (2018) and Kovacev-Nikolic et al. (2016). However, we were not able to set them up in a way that led to competitive results. We suspect that the reason for this was the great number of features (around 500,000 per line of sight) in our persistence diagrams. The problem we were facing was that because of this number of features, we obtained a very large and noisy data vector. We reduced its dimension using a principal component analysis, similar to Kovacev-Nikolic et al. (2016), but unfortunately, the quality of the resulting data was not good enough to obtain sufficiently tight bounds on the cosmological parameters. This might change when the principal component analysis can be applied to a less noisy data vector that is extracted from a larger training sample.

Heatmaps are the method that worked best for us in the present setting. These are defined in the spirit of the multi-scale kernel introduced in Reininghaus et al. (2015). The idea is to replace each point in a persistence diagram with a Gaussian. More precisely, one considers the diagram as a discrete measure (i.e. a sum of Dirac delta distributions) on $D \subset \mathbb{R}^2$, where $D = \{(x, y) \in \mathbb{R}^2 | x < y\}$ and convolves this with a two-dimensional isotropic Gaussian distribution. The result is for every value $t > 0$ a continuous function $u_t(x, y) : D \to \mathbb{R}$ that can be seen as a smoothed version of the persistence diagram. The value t is called the scaling parameter and determines how much smoothing is applied to the initial diagram. For an example of such a heatmap, see Fig. 3.

We compute the heatmaps in the following way: we first compute the persistence diagram Dgm^M for the S/N maps of each tiled realisation of the DES-Y1 footprint and for each tomographic bin. For each of these, we create a two-dimensional histogram of the persistence diagram with 100×100 bins. For Dgm_0 our bins cover the S/N range $[-3, 2]^2$, for Dgm_1 the bins cover the S/N range $[-1, 4]^2$. The upper limit of 4 in the S/N maps avoids issues with source-lens coupling as elaborated in Martinet et al. (2018). All persistence features that lie outside of this range are projected to the edge of the respective bin ranges. After-

wards, we convolve these histograms with a Gaussian kernel of scaling parameter $t = 0.2$ using two-dimensional FFTs[7].

3.3. Data Compression

To perform a Bayesian cosmological parameter inference, we compress the data provided by the persistence heatmaps. We explored several compression methods, which are discussed in App. A. In the end, we opted for an adaptation of our method developed in H+21; we iteratively build a data vector in the following way: As a first step, for each pixel x of a heatmap, we compute the mean squared difference between the single cosmologies of cosmo-SLICS and their mean, weighted by the inverse variance within the SLICS

$$\Delta x_{\mathrm{weighted}} \equiv \sum_{i=0}^{25} \frac{\left(x_{\mathrm{cosmoSLICS},i} - \langle x_{\mathrm{cosmoSLICS}}\rangle\right)^2}{\sigma^2(x_{\mathrm{SLICS}})} \,. \tag{6}$$

This $\Delta x_{\mathrm{weighted}}$ describes the cosmological information content of a pixel from the heatmap, as it quantifies how much its value varies between different cosmologies with respect to the expected standard deviation. As the first point of our data vector, we choose the one with the highest cosmological information content. Then we proceed to add more points in the following way: assuming we already have n entries in our data vector, we determine the next entry from the mean squared difference, weighted by the inverse sub-covariance matrix estimated from the SLICS. In other words: Let $\Delta\boldsymbol{x}_i \equiv \boldsymbol{x}_{\mathrm{cosmoSLICS},i} - \langle \boldsymbol{x}_{\mathrm{cosmoSLICS}}\rangle$ be the difference between the data vector of the i-th cosmology of cosmo-SLICS and the mean data vector of all cosmo-SLICS. For each pixel in the heatmap that is not already part of the data vector $\boldsymbol{x}$, we create a new data vector $\boldsymbol{x}'$ that contains this pixel, and then we compute

$$\hat{\chi}^2 = \sum_{i=0}^{25} \Delta\boldsymbol{x}_i' C^{-1}_{\mathrm{SLICS}} \Delta\boldsymbol{x}_i' \,. \tag{7}$$

The pixel yielding the highest $\hat{\chi}^2$ is then added to the data vector, and the procedure is repeated until we have reached the desired amount of data points. Again, this serves to maximise the cosmological information content of our data vector with respect to the expected covariance. To ensure that our data vector follows a Gaussian distribution, we only consider elements of the heatmaps that count at least 100 features (compare Fig. A.2). We found that 12 data points per tomographic bin combination yield good results, but the dependence on the number of data points is weak. An example of such a resulting data vector can be seen in Fig. 4.

While this method certainly does not capture all of the information residing in the heatmaps, this 'χ^2-maximiser' manages to capture most information and is therefore competitive with the other data compression methods. A comparison is given in App. A.

[7] We tried different values between $t = 0.05$ and $t = 0.4$. The results were stable with respect to these changes, and the value of 0.2 appears to be a good compromise between stability and precision.

3.4. Two-point statistics

The established methods to infer statistical properties of the matter and galaxy distribution concentrate on the second-order statistics such as the 2PCFs, their Fourier counterparts, the power spectra, or derived measures such as COSEBIs (Schneider et al. 2010; Asgari et al. 2020). The key advantage of these statistics over others is that, although they capture only the Gaussian information of the large-scale structure, they can be calculated analytically from the well-understood matter power spectrum $P(k, z)$. Indeed, the lensing power spectrum between galaxies of tomographic bin i with redshift distribution $n^i(z)$ and those of tomographic bin j with $n^j(z)$ is modelled in the Limber approximation as

$$C_\ell^{ij} = \int_0^{\chi_{\mathrm{H}}} \frac{W^i(\chi) W^j(\chi)}{\chi^2} P\left(\frac{\ell + 1/2}{\chi}, z[\chi]\right) \mathrm{d}\chi \,, \tag{8}$$

where χ_{H} is the co-moving distance to the horizon and $W(\chi)$ is the lensing efficiency defined as

$$W(\chi) = \frac{3\Omega_{\mathrm{m}} H_0^2}{2c^2} \int_\chi^\infty \mathrm{d}\chi' \, \frac{\chi(\chi' - \chi)}{\chi' a(\chi)} \, q(\chi') \,. \tag{9}$$

Here, $q(\chi) = n(z[\chi]) \frac{\mathrm{d}z[\chi]}{\mathrm{d}\chi}$ is the line of sight probability density of the galaxies, H_0 the Hubble parameter and c the speed of light. From the projected lensing power spectrum, the cosmic shear correlation functions $\xi_\pm^{ij}$ are computed as

$$\xi_\pm^{ij}(\vartheta) = \frac{1}{2\pi} \int_0^\infty C_\ell^{ij} J_{0,4}(\ell\vartheta) \, \ell \, \mathrm{d}\ell \tag{10}$$

where $J_{0,4}$ are the Bessel functions of the first kind. To compute the theoretical two-point correlation functions we calculate the power spectrum $P(k)$ using the public HALOFIT model (Takahashi et al. 2012).

We use the software TREECORR (Jarvis et al. 2004) to estimate the 2PCF $\hat{\xi}_\pm^{ij}(\vartheta)$ from the simulations and the DES-Y1 lensing data, computed as

$$\hat{\xi}_\pm^{ij}(\vartheta) = \frac{\sum_{a,b} w_a w_b \left[\epsilon_{\mathrm{t}}^i(\boldsymbol{\theta}_a)\epsilon_{\mathrm{t}}^j(\boldsymbol{\theta}_b) \pm \epsilon_\times^i(\boldsymbol{\theta}_a)\epsilon_\times^j(\boldsymbol{\theta}_b)\right]}{\sum_{a,b} w_a w_b S_a S_b} \,, \tag{11}$$

where the sums are over all galaxy pairs (a, b) in tomographic bins (i, j) that are inside the corresponding ϑ-bin. As in HD+21, we used 32 logarithmically spaced ϑ-bins in the range $[0.5', 475'.5]$, although not all angular scales are used in the parameter estimation (see the following section).

3.5. Cosmological parameter estimation

As in H+21, we train a GPR emulator using data extracted from the 26 different cosmo-SLICS models to interpolate our data vector at arbitrary cosmological parameters within the training range. We refer the reader to H+21 and HD+21 for more details on the emulator. We assess its accuracy by performing a leave-one-out cross-validation: we remove one cosmology of the cosmo-SLICS from our training sample and let the GPR-emulator predict this cosmology, training on the other 25. We repeat this procedure for all 26 cosmologies and use the mean squared difference between predictions and truth as an estimate for the error of the

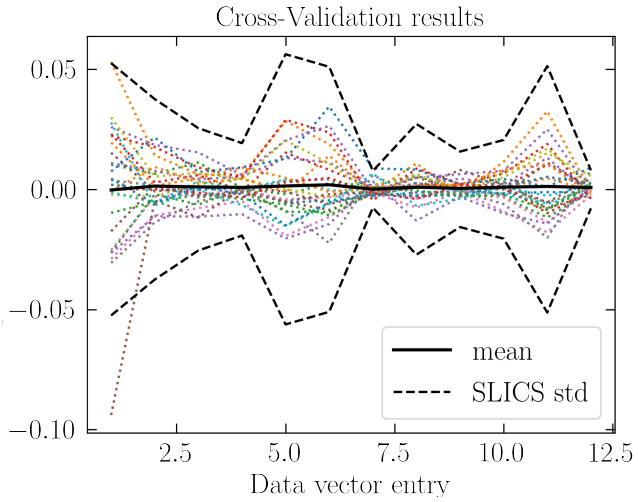


Fig. 5. Accuracy of the GPR emulator evaluated by a leave-one-out cross-validation, shown here for the case where the aperture mass maps are constructed from the concatenation of all four tomographic redshift bins (i.e. no tomography). The x axis depicts the data vector entry, and the y axis the relative difference between predicted and measured value. The 26 individual dotted lines correspond to one cosmology that is left out of the training set and then predicted, the solid black line is the mean of all dotted lines. The black dashed lines depict the standard deviation from the Covariance Training Set.

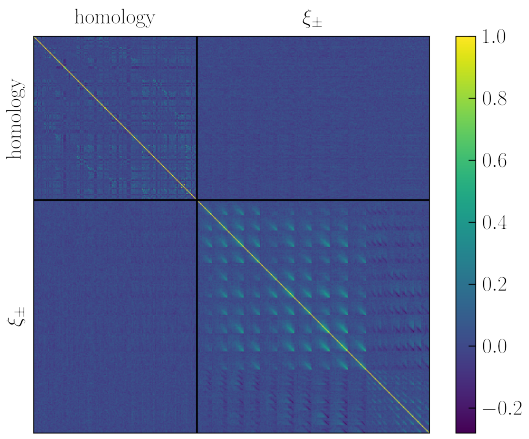


Fig. 6. Correlation matrix for a joint analysis with both persistent homology and two-point correlation functions.

emulator, which is typically well below the statistical error (compare Fig. 5). We then add this to the diagonal of our sample covariance matrix to account for uncertainties in the modelling.

An alternative method to estimate the uncertainty of the predictions is to use the error provided by the GPR emulator itself. We also tested this method and found that, while this method is a bit slower (since the inverse covariance matrix needs to be re-computed in every step of the MCMC), it provides comparable, albeit slightly tighter constraints than the first method. In the end, we opted for the more conservative choice of estimating the modelling uncertainties via cross-validation.

As in HD+21, we then integrate our GPR emulator into the CosmoSIS analysis pipeline (Zuntz et al. 2015) and infer the cosmological parameters by sampling the likelihood using the polychord sampling method (Handley et al. 2015), which constitutes a good compromise between speed and accuracy (Lemos et al. 2022). A few relatively minor changes to the CosmoSIS likelihood module allow for an easy and fast joint analysis of both persistent homology statistics and shear two-point correlation functions.

Finally, we estimate our sample covariance matrix from the 124 survey realisations of the Covariance Training Set. Specifically, we compute a matrix for each of the 10 different realisations of the shape noise and use the average over those 10 covariance matrices as our best estimate. Here, we randomly distribute the 124 lines of sight for the 19 regions to avoid over-estimating the sample variance (compare HD+21). Further, since the inverse of a simulation-based covariance matrix is generally biased (Hartlap et al. 2007), we mitigate this effect by adopting a multivariate t-distribution likelihood (Sellentin & Heavens 2016). The extracted covariance matrix can be seen in Fig. 6, the priors used for cosmological parameter estimation are listed in Tab. 1.

Table 1. Prior ranges of cosmological and nuisance parameters in the likelihood analysis.

Parameter	Prior Type	Prior range
Ω_m	uniform	$[0.1, 0.55]$
h	uniform	$[0.6, 0.82]$
Ω_b	delta	0.0447
τ	delta	0.08
n_s	delta	0.969
σ_8	uniform	$[0.53, 1.3]$
w_0	uniform	$[-2.0, -0.5]$
w_a	delta	0
S_8	uniform	$[0.6, 0.9]$
A_IA	uniform	$[-5, 5]$
baryon feedback	uniform	$[0, 2]$
Δm_1	Gaussian	$\mu = 0.12,\ \sigma = 0.023$
Δm_2	Gaussian	$\mu = 0.12,\ \sigma = 0.023$
Δm_3	Gaussian	$\mu = 0.12,\ \sigma = 0.023$
Δm_4	Gaussian	$\mu = 0.12,\ \sigma = 0.023$
$\mathrm{d}z_1$	Gaussian	$\mu = 0,\ \sigma = 0.008$
$\mathrm{d}z_2$	Gaussian	$\mu = 0,\ \sigma = 0.014$
$\mathrm{d}z_3$	Gaussian	$\mu = 0,\ \sigma = 0.011$
$\mathrm{d}z_4$	Gaussian	$\mu = 0,\ \sigma = 0.009$

Notes. Priors on cosmological parameters are provided by the range of our Cosmology Training Set, prior on Δm_i and on A_IA are from T+18, while those on $\mathrm{d}z_i$ are from Joudaki et al. (2020).

4. Mitigating systematic effects

Our cosmological parameter analysis needs to account for systematic effects that are known to affect cosmic shear data. The most important ones for this work are intrinsic alignments of source galaxies, baryonic physics, multiplicative shear bias and uncertainties in the redshift estimation of galaxies (Mandelbaum 2018). On top of these, limits in the force resolution of the cosmo-SLICS might introduce a bias into our modelling, plus source clustering can produce systematic differences between the data and the simulations, in which the latter is absent. While we investigate the

former, the latter has been shown in HD+21 to be largely subdominant in the aperture mass statistics measured in the DES-Y1 data and is therefore neglected here. In this section, we explain how these systematic effects affect our 2PCFs and persistent homology measurements and detail the mitigation strategies we chose to account for their impact.

4.1. 2PCF

We use the public modules in COSMOSIS to marginalise over the impact of intrinsic alignments. Following T+18, we model IA with the non-linear alignment model (Bridle & King 2007, hereafter referred to as the NLA model), which adds a contribution to the matter power spectrum that propagates into the lensing signal following Eqs. (8) and (10). More sophisticated IA models, including tidal torque terms (notably the Tidal Alignment and Tidal Torque model, or TATT Blazek et al. 2019) have been used recently in cosmic shear analyses, but there is no clear evidence that the data prefer such a model over the NLA (T+18, Secco et al. 2022). The NLA model can have multiple parameters (amplitude, redshift dependence, luminosity dependence, pivot scales, colour); however, we follow HD+21 and vary only the amplitude ($A_{\rm IA}$) and luminosity (α) parameters, considering no other dependencies. This is justified by the weak constraints that exist on them in the DES-Y1 data (compare T+18). The parameters $A_{\rm IA}$ and α are allowed to vary in the range [-5.0, 5.0].

Following the fiducial DES-Y1 choices, the impact of baryon feedback is minimised by cutting out angular scales in the $\xi_\pm$ statistics where unmodelled baryonic physics with a strong AGN model[8] could shift the data by more than 2%. We, therefore, exclude from our analysis the same small scales as those of T+18, which are different for ξ_+ and ξ_-, and further vary with redshift.

The shear inference is obtained with the METACALIBRATION method in this work, which has a small uncertainty that can be captured by a shape calibration factor Δm, which multiplies the observed ellipticities as $\epsilon_{1/2} \rightarrow \epsilon_{1/2}(1+\Delta m)$. As described in T+18, Δm is a nuisance parameter that we sample by a Gaussian distribution with a width of 0.023, centred on 0.012 when analysing the data and on zero when analysing simulations. COSMOSIS includes this nuisance on the two-point function model directly, namely $\xi_\pm^{ij} \rightarrow \xi_\pm^{ij}(1+\Delta m^i)(1+\Delta m^j)$. The priors on m are listed in Tab. 1.

Photometric errors in the 2PCFs are mitigated by using the generic module within COSMOSIS, which shifts the $n^i(z)$ by small bias parameters Δz^i and updates accordingly the lensing predictions. These bias parameters are sampled from Gaussian distributions with widths corresponding to the posterior DIR estimates of the mean redshift per tomographic bin 'i', also tabulated in Tab. 1.

4.2. Persistent homology

As mentioned in Sec. 2.2, we assess the impact of systematics on the topology of aperture mass maps by using the Systematics Training Set, which are numerical simulations specifically tailored for this exercise. Following HD+21, we neglect the cosmology scaling of these systematics and only evaluate their relative impact at the fiducial cosmology. We find that the overall impact of systematic effects is sufficiently well captured by a linear modelling strategy: for each systematic effect with respective nuisance parameter λ (i.e. $A_{\rm IA}$ for intrinsic alignments, Δz for redshift uncertainties, Δm for multiplicative bias and $b_{\rm bar}$ for baryons), we measure the impact $\boldsymbol{x}_{\rm sys}(\lambda)$ on the measured data vector from the associated Systematics Training Set and fit each point of the data vector with a straight line:

$$\boldsymbol{x}_{\rm sys}(\lambda) = \boldsymbol{m}_{\boldsymbol{x}}\lambda + \boldsymbol{x}_{\rm nosys} \ . \tag{12}$$

In particular, $\boldsymbol{x}(0) \equiv \boldsymbol{x}_{\rm nosys}$ is the data vector which is not impacted by any systematic effects. For a given set of values of the nuisance parameters, we combine these different sources of uncertainty to model our systematics-infused data vector as:

$$\boldsymbol{x}_{\rm sys} = \boldsymbol{x}_{\rm nosys} + \boldsymbol{m}_{\rm IA}A_{\rm IA} + \boldsymbol{m}_{\rm bar}b_{\rm bar} + \boldsymbol{m}_{\rm dz}\Delta z + \boldsymbol{m}_{\Delta \rm m}\Delta m \ . \tag{13}$$

While this certainly constitutes a simplified approach that does not capture potential cross-correlations between different systematic effects nor any cosmology dependence, we consider it sufficient at the current level of uncertainties (compare Fig. 7). To compute $\boldsymbol{m}_{\rm IA}A_{\rm IA}$, the Intrinsic Alignments mocks are infused with $A_{\rm IA}$ values of [-5.0, -2.0, -1.0, 0.5, 0.0, 0.5, 1.0, 2.0 and 5.0], however we set the redshift dependence to zero, given the weakness of the constraints on this parameter in the DES-Y1 data (see T+18). In the upper left panel of Fig. 7 we report the fractional effect on the signal and observe that positive IA suppresses the elements of the data vector. This is caused by the partial cancellation of the lensing signal by IA, which attenuates the contrasts in the aperture mass maps, which translates into a topological structure with fewer features. The figure also presents the results as modelled by the linear interpolation, which reproduces the nodes on which the training was performed to sufficient accuracy, indicating that our approach is adequate to model IA, at least for the range of $A_{\rm IA}$ values tested here.

The impact of shear calibration uncertainty is modelled by measuring the statistics for ellipticities modified with four values of Δm, namely -0.025, -0.0125, 0.0125, and 0.025, and once again fitting a straight line through each element of the homology data vector as a function of Δm. The results are presented in the upper right panel of Fig. 7, showing that within this range, shear calibration affects the statistics by less than one per cent except for two elements, which are affected by up to 4%.

Photometric uncertainties are modelled from the dedicated Systematics Training Set in which the $n(z)$ in each tomographic bin has been shifted by 10 values Δz^i, from which we are once again able to fit a linear response for each element of the data vector. In the case of cross-redshifts, the mean of all shifts is used to compute the derivative, as in HD+21. The lower right panel of Fig. 7 shows the impact on the data vector, which is sub-dominant compared to the IA. This is largely due to the tight priors on Δz^i that we are able to achieve with the DIR method, as derived in Joudaki et al. (2020) and reported in Tab. 1.

The Magneticum simulations are used in a similar way to test the impact of baryonic feedback, with the main difference that we can only fit $\boldsymbol{m}_{\rm bar}$ on two points: the simu-

[8] The power spectrum of the OWLS AGN model (van Daalen et al. 2014) is used for this assessment.

lations with and without baryons ($b_{\rm bar}$ = 1 and 0, respectively). We nevertheless apply the same methodology here, which allows us to interpolate between these two cases to mimic milder models (e.g. $b_{\rm bar}$ = 0.5) and even to extrapolate and explore stronger feedback models ($b_{\rm bar}$ >1.0). The lower right panel of Fig. 7 shows that baryonic feedback with $b_{\rm bar}$ = 1.0 has almost as much importance as an IA model with $A_{\rm IA}$= 1.0 and should therefore not be neglected in this analysis.

The last systematic effect that we include in this analysis is the impact of the force accuracy in the N-body simulations that are used in the modelling. We inspect the difference between the data vector measured from the high-accuracy mocks to that of the main Cosmology Training Sample at the fiducial cosmology and find that the overall impact of this effect is sub-dominant to the sample variance of the SLICS. Nevertheless, we measure the ratio of the high-resolution data vector over the fiducial one and apply it as a correction factor to re-calibrate our model in the analysis of observed data.

4.3. Mitigation strategy

We further estimate the impact of the different systematic effects on the cosmology inference by running likelihood analyses for data vectors that have been infused with one systematic effect while keeping these unmodelled. The results of these tests can be seen in Fig. 8. We observe that the baryons have a small impact on the inferred $\Omega_{\rm m}$ and can bias S_8 by 1σ, assuming $b_{\rm bar}$=1.0. Unmodelled IA (with $A_{\rm IA}$=1.0) tend to bias both $\Omega_{\rm m}$ and S_8 towards lower values; both photometric redshift uncertainties and multiplicative shear bias have a minor impact on the posterior constraints, given the tight priors available on Δz and Δm.

We finally investigate how marginalisation over the different systematic biases changes the posterior contours in our likelihood analysis in Fig. 9. We find that marginalisation over baryonic effects and intrinsic alignments both decrease the constraining power on $\Omega_{\rm m}$ and S_8 by about 25%, whereas the marginalisation over multiplicative shear biases and photometric redshift uncertainties have a negligible impact. Both analyses suggest that the impact of systematic effects on persistent homology statistics is noticeable but not severe and that our marginalisation strategies work as expected.

5. Validation

In this work, we want to investigate whether a likelihood analysis of tomographic cosmic shear data with persistent homology is feasible and whether a joint analysis with two-point statistics yields more information than an analysis that solely utilises two-point statistics. For this purpose, we perform three likelihood analyses of the same mock data extracted from the Covariance Training Set: one solely with two-point correlation functions that we model within the COSMOSIS pipeline, one solely with our persistent homology method, and finally the combined analysis.

As can be seen in Fig. 10 and Tab. 2, the persistent homology analysis is already able to constrain S_8 better than the two-point analysis ($S_8 = 0.817^{+0.040}_{-0.028}$ for persistent homology versus $S_8 = 0.772 \pm 0.043$ for two-point statistics). However, a joint analysis offers several additional benefits. While two-point statistics are able to constrain the parameter to $A_{\rm IA} = -0.19^{+0.90}_{-0.40}$, persistent homology yields $A = 0.47^{+0.64}_{-0.56}$ and a joint analysis is able to reduce the error bars to $A_{\rm IA} = 0.29 \pm 0.36$. Apart from tighter constraints on S_8 ($S_8 = 0.815^{+0.030}_{-0.021}$ for a joint analysis), a joint analysis also yields competitive lower limits on the equation-of-state parameter of dark energy ($w_0 > -1.14$ at 68% confidence), while two-point statistics are unable to place any constraints on this parameter, with our choice of sampling method. Most importantly, all cosmological and nuisance parameters are recovered within 1σ. We thus conclude that our analysis pipeline has been validated and move on towards a cosmological parameter analysis of real data.

6. Results

Having validated our analysis pipeline, we now use it to perform our cosmological parameter analyses using the DES-Y1 data. In order to do that, we split the source galaxy catalogue into the same 19 tiles as our mock data and compute the persistence statistics as well as the two-point correlation functions for each tile individually.

The results can be seen in Fig. 11. We observe that neither persistent homology nor two-point correlation functions are able to place meaningful constraints on the equation of state parameter for dark energy, w_0. For the matter clustering parameter S_8, the constraints from persistent homology ($S_8 = 0.747^{+0.025}_{-0.031}$) are tighter than, but fully consistent with, the constraints from two-point correlation functions ($S_8 = 0.759^{+0.049}_{-0.042}$). The same goes for the amplitude of galaxy intrinsic alignments ($A = 1.54 \pm 0.52$ for persistent homology and $A = 1.33^{+0.92}_{-0.56}$ for two-point correlation functions). In particular, this implies that persistent homology detects the intrinsic alignment effect roughly at the 3σ level. Interestingly, the constraints for the matter density parameters are not consistent ($\Omega_{\rm m} = 0.468^{+0.051}_{-0.036}$ for persistent homology and $\Omega_{\rm m} = 0.256^{+0.034}_{-0.058}$ for two-point correlation functions). Hamana et al. (2020) observed a similar trend when observing data from the HSC: While a real- and Fourier-space analysis yield perfectly consistent values for S_8, a slight tension between the $\Omega_{\rm m}$ constraints can be observed in their Fig. 15. We observe a much larger tension that prevents us from performing a joint parameter analysis, which would tighten the S_8-constraints considerably. We discuss this in more detail in App. B, where we show that such a tension arises in about 0.5% of all cases due to a mere statistical fluctuation. A visual inspection of the data vector suggests that this tension might be caused by the highest signal-to-noise peaks (see App. B), but when excluding these in a parameter analysis, we find only a marginal improvement of the tension that is likely just due to the loss of constraining power. We note that the upper limit in the constraints for $\Omega_{\rm m}$ from persistent homology barely passes our criterion not to be dominated by the prior (compare Tab. 2). This means that a wider prior for $\Omega_{\rm m}$ would likely lead to a higher upper limit in the constraints. The lower limit should not be strongly affected by the prior, as the likelihood of $\Omega_{\rm m}$ already started falling for $\Omega_{\rm m} > 0.5$.

Comparing our results with the ones from peak count statistics, where HD+21 measured $S_8 = 0.737^{+0.027}_{-0.031}$ on the same data set, we observe remarkably consistent results (compare Fig. 11). The trend towards high values

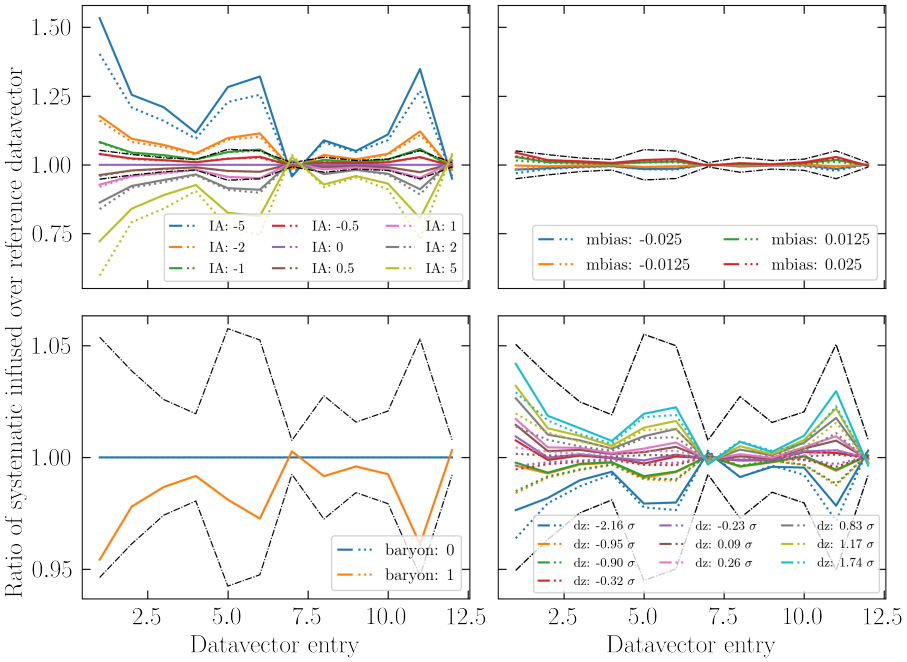


Fig. 7. Impact of the main systematic effects on the data vector. For each systematic, we show the measured (solid line) and interpolated (dotted line) ratio of the systematic-infused data vector over a reference data vector. For simplicity, we only show the results for the combination of all four tomographic redshift bins. The black dashed lines correspond to the 1σ standard deviation estimated from the Covariance Training Set.

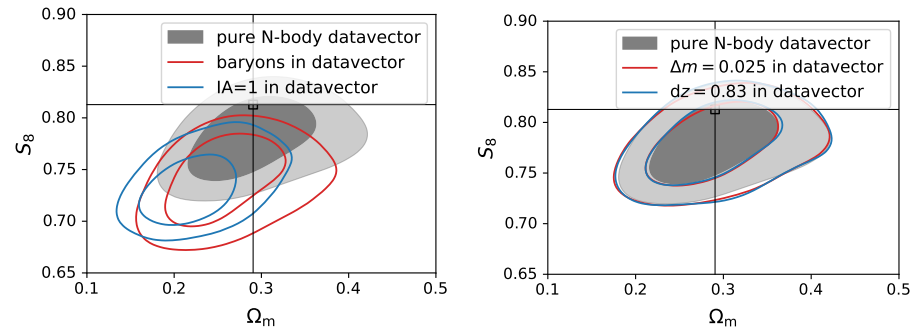


Fig. 8. Impact of unmodelled systematic biases on the posterior of a likelihood analysis with heatmaps. In all cases, we do not marginalise over any systematic effects. The target data vector is then infused with one systematic bias, and we run a likelihood analysis for this infused data vector. For comparison, we show the constraints on a data vector that is not infused by systematics (grey). Note that the values of the dz shifts are given in units of the standard deviation of the dz prior (compare Tab. 1).

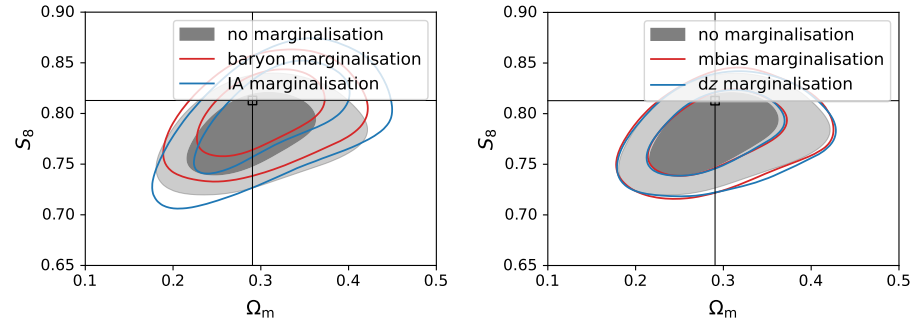


Fig. 9. Effects of marginalising over different systematic effects. In all cases, we perform a likelihood analysis on mock data, marginalising over one systematic effect. For comparison, we show the constraints we achieve when we do not marginalise over any systematics (grey). The case where we marginalise over all systematics corresponds to the blue contours in Fig. 10

.

of Ω_m can also be observed in HD+21 (see in particular Fig. 17 and 18 in HD+21). This is particularly interesting since their constraints have been achieved using a pipeline that is fully independent of ours, utilising a different statistic on signal-to-noise maps of aperture masses constructed with an independent code, albeit based on the same set of N-body simulations. Furthermore, we can see that the constraints achieved from persistent homology outperform the ones from peak statistics, albeit not by much. This improvement is still significant since, contrary to HD+21, we include an error estimate for the emulator and a marginalisation over intrinsic alignments and baryonic effects, which decrease the constraining power of our analysis pipeline.

Comparing our results to T+18, we see that our results from two-point statistics are a bit different ($S_8 = 0.777^{+0.036}_{-0.038}$ in T+18 and $S_8 = 0.759^{+0.049}_{-0.042}$ here), which is driven mainly by the different redshift distribution estimates (as shown in Joudaki et al. 2020). Considering the intrinsic alignment effect, we achieve consistent, but tighter constraints (compare the NLA case of Fig. 16 in T+18). Regarding the tension we measure for Ω_m, T+18 report $\Omega_m = 0.274^{+0.073}_{-0.042}$, which is fully consistent with our results

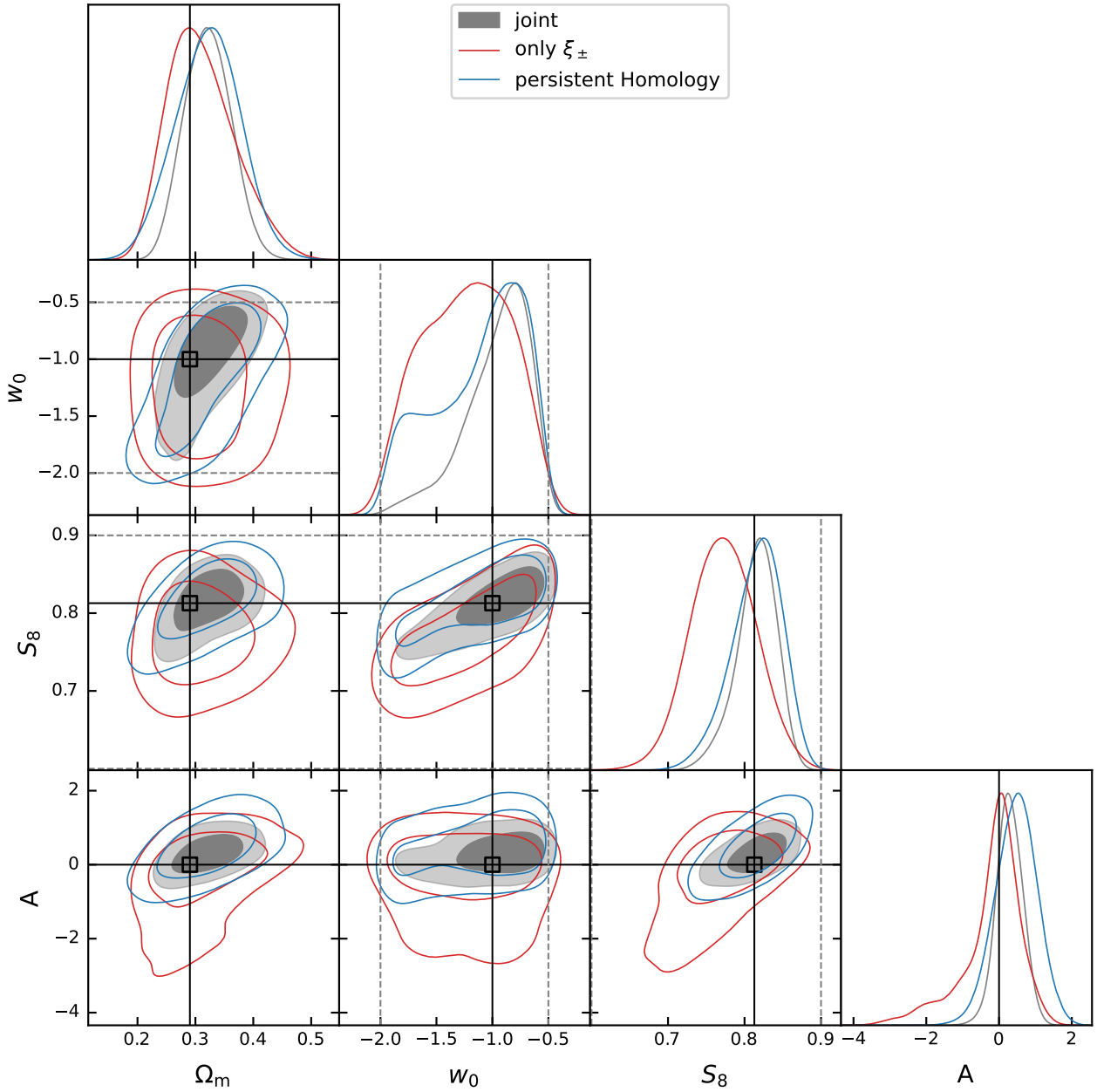


Fig. 10. Results of likelihood analyses for DES-Y1 mock data. We show the results for two-point statistics (red), persistent homology (blue), and for the joint analysis (grey, filled). The dotted lines show the prior ranges, the solid black lines visualise the true value of each parameter, and the black crosses denote the nodes of our Cosmology Training Set. The complete results can be seen in Fig. C.2, the marginalised posterior constraints can be seen in Tab. 2.

Table 2. Posterior 68% confidence intervals on cosmological and nuisance parameters from the likelihood analyses in Fig. 10 and Fig. 11.

Method	Ω_m	S_8	w_0	A
Validation (mock data)				
persistent homology	$0.323^{+0.059}_{-0.053}$	$0.817^{+0.040}_{-0.028}$	–	$0.47^{+0.64}_{-0.56}$
$\xi_\pm$	$0.311^{+0.046}_{-0.069}$	0.772 ± 0.043	–	$-0.19^{+0.90}_{-0.40}$
joint	0.321 ± 0.040	$0.815^{+0.030}_{-0.021}$	> -1.14	0.29 ± 0.36
DES-Y1 data				
persistent homology	$0.468^{+0.051}_{-0.036}$	$0.747^{+0.025}_{-0.031}$	< -1.04	1.54 ± 0.52
$\xi_\pm$	$0.256^{+0.034}_{-0.058}$	$0.759^{+0.049}_{-0.042}$	> -1.47	$1.33^{+0.92}_{-0.56}$

Notes. Constraints are only cited if the value of the marginalised posterior does not surpass 13.5% at the edge of the priors (Asgari et al. 2021).

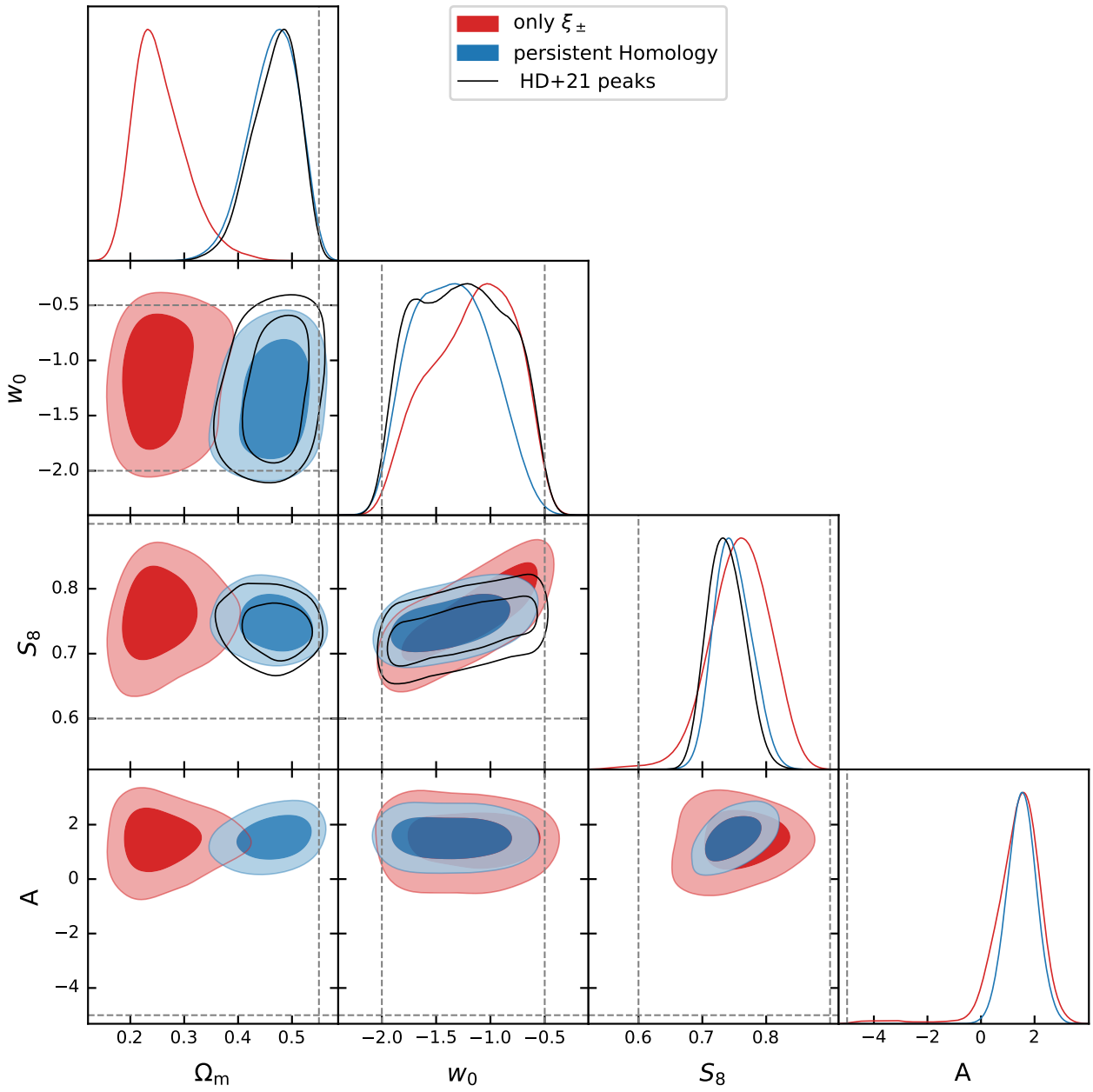


Fig. 11. Results of our likelihood analyses for the DES-Y1 survey. We show the results for two-point statistics (red), persistent homology (blue) and report as well the constraints achieved by HD+21 with peak count statistics (black). The dotted lines denote the prior ranges, the black crosses denote the nodes of our Cosmology Training Set. The complete results including nuisance parameters can be seen in Fig. C.3, while the marginalised posterior constraints can be seen in Tab. 2.

from two-point correlation functions and also disagrees with our constraints from persistent homology.

7. Discussion

In this work, we carried out a likelihood analysis on tomographic cosmic shear data using persistent homology, including the marginalisation over systematic effects, and have shown from simulated data that the posterior constraints can be significantly improved in a joint analysis. While this holds true especially for the intrinsic alignment parameter $A_{\rm IA}$ and the equation of state of dark energy w_0, the constraints on the matter clustering parameter S_8 also improve substantially.

For our analysis, we had to make a number of choices, including which persistence statistic to use, which smoothing scale to apply to the heatmaps, and which data compression method to utilise. We have noticed that the posterior constraints achieved by a likelihood analysis do not strongly depend on any of these choices, as can be seen for example in Fig. C.1. While further fine-tuning could probably slightly improve the constraining power of our analysis, we believe that this overall stability with respect to different analysis choices provides strong evidence that we have reached the true sensitivity of persistence statistics to cosmology.

When applying our analysis pipeline to real data, we find that high values of the matter density parameter $\Omega_{\rm m}$ are preferred, as observed in HD+21, where a fully inde-

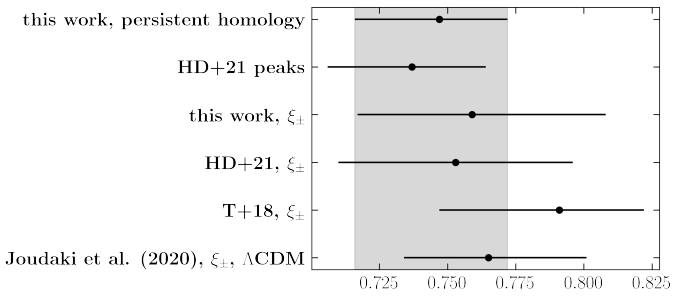


Fig. 12. Comparison of the constraints on the matter clustering parameter S_8 from DES-Y1 survey data in a wCDM cosmology with fixed neutrino mass.

pendent pipeline and a different summary statistic were utilised. The remarkable similarity between these results suggests that peak count statistics and persistent homology quantify similar aspects of the large-scale structure distribution. The fact that both methods favour larger values of Ω_m may point to a statistical fluctuation in the DES-Y1 data or an unknown effect modifying the topological structure of the matter distribution. We investigate this in App. B and show that the chance of such a tension arising due to a statistical fluctuation in the data is about 0.5%. We also note that our underlying training data, the shear catalogues of the Cosmology Training Sets, are the same as the ones used in HD+21, so this bias might also point towards a statistical fluctuation in this training set. A larger simulation suite would be able to shine a light on this, even though this seems unlikely given the fact that the tension does not exist when validating on simulated data (See Fig. 10). Our analysis of the tension in App. B, a visual inspection of the data suggests that the tension might be at least partly driven by the high signal-to-noise peaks in the aperture mass maps, which carry both the most cosmological signal and are most affected by systematics. However, excluding those peaks did not remove the tension, which means that it can certainly not be fully explained by these high signal-to-noise peaks.

We performed several consistency checks to investigate whether the tension is artificially created by our analysis setup: We tried a parameter inference with two-point correlation functions by measuring them in the Cosmology Training Set and emulating them via the same pipeline that we used for persistent homology. Furthermore, we tried removing some nodes from the Cosmology Training Set and applying different methods of data compression. The results were stable under all these tests, suggesting that the simulation-based inference is not the driver of the tension we measure for Ω_m.

Another possible explanation is that the DES-Y1 data include an effect that we have not accounted for. This might be an unknown systematic or a sign of new physics. For example, 2PCF are not sensitive to primordial non-Gaussianities, whereas persistent homology is (Biagetti et al. 2022).

Overall our constraints on S_8 are consistent with previous works (see Fig. 12); the largest discrepancy is between our analysis and the one from T+18, which is mainly driven by a different method of estimating the source redshift distribution. When comparing our results to similar works with peak count statistics, the constraining power of persistent homology appears to be slightly better. In addition, there are a few key differences between our work and HD+21. Firstly, while HD+21 apply a boost factor to account for baryons, we marginalise over continuous baryonic effects (and intrinsic alignments) with a wide prior, which is inflating our constraints. Secondly, we account for the emulator uncertainty as described in Sect. 3.5; this is not done in HD+21. Comparing with H+21, we see that this emulator uncertainty also inflates cosmological parameter constraints, indicating that the contours reported in HD+21 may be slightly too small. This effect is amplified by the fact that we were only able to train our emulator on 9 lines of sight per cosmology, compared to the 50 lines of sight in H+21. Lastly, and most importantly, we have shown in H+21 that persistent homology excels in a high signal-to-noise range, which is not accessible in a tomographic analysis of current-generation surveys. We thus expect this method to out-perform several other higher-order statistics in next-generation surveys.

Acknowledgements. SH acknowledges support from the German Research Foundation (DFG SCHN 342/13), the International Max-Planck Research School (IMPRS) and the German Academic Scholarship Foundation. SU acknowledges support from the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck-Humboldt Research Award endowed by the Federal Ministry of Education and Research. Joachim Harnois-Déraps acknowledges support from an STFC Ernest Rutherford Fellowship (project reference ST/S004858/1). Computations for the N-body simulations were enabled by Compute Ontario (www.computeontario.ca), Westgrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). The SLICS numerical simulations can be found at http://slics.roe.ac.uk/, while the cosmo-SLICS can be made available upon request. TC is supported by the INFN INDARK PD51 grant and by the fare Miur grant 'ClustersXEuclid' R165SBKTMA. KD acknowledges support by the COMPLEX project from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement ERC-2019-AdG 882679 the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 - 390783311 and by the DFG project nr. 490702358. The Magneticum Simulations were carried out at the Leibniz Supercomputer Center (LRZ) under the project pr86re and pr83li. We thank the anonymous referee for insightful comments given on a previous version of this article.

Author contributions: All authors contributed to the development and writing of this paper. After the lead author, they are separated into two groups, both listed alphabetically. In the first group, BB developed the necessary mathematical background, PB oversaw the measurement and validation of two-point statistics, JHD provided the suites of numerical simulation tailored to the measurement and SU developed the integration into COSMOSIS. In the second group, TC, KD and NM were responsible for running and post-processing the Magneticum simulations, extracting the mass maps and computing the lensing statistics.

References

Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, ApJS, 239, 18
Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, Phys. Rev. D, 105, 023520
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, A&A, 645, A104
Asgari, M. & Schneider, P. 2015, A&A, 578, A50
Asgari, M., Tröster, T., Heymans, C., et al. 2020, A&A, 634, A127
Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291
Benítez, N. 2000, ApJ, 536, 571
Bergé, J., Amara, A., & Réfrégier, A. 2010, ApJ, 712, 992
Biagetti, M., Calles, J., Castiblanco, L., Cole, A., & Noreña, J. 2022, arXiv e-prints, arXiv:2203.08262
Biagetti, M., Cole, A., & Shiu, G. 2020, arXiv e-prints, arXiv:2009.04819

Biffi, V., Dolag, K., & Böhringer, H. 2013, MNRAS, 428, 1395
Blazek, J. A., MacCrann, N., Troxel, M. A., & Fang, X. 2019, Phys. Rev. D, 100, 103506
Bocquet, S., Saro, A., Dolag, K., & Mohr, J. J. 2016, MNRAS, 456, 2361
Bridle, S. & King, L. 2007, New Journal of Physics, 9, 444
Bubenik, P. 2015, Journal of Machine Learning Research (JMLR), 16, 77
Burger, P., Friedrich, O., Harnois-Déraps, J., & Schneider, P. 2022, A&A, 661, A137
Castro, T., Borgani, S., Dolag, K., et al. 2021, MNRAS, 500, 2316
Castro, T., Quartin, M., Giocoli, C., Borgani, S., & Dolag, K. 2018, MNRAS, 478, 1305
Chazal, F. & Michel, B. 2021, Frontiers in Artificial Intelligence, 4, 108
Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, MNRAS, 499, 5902
Chittajallu, D. R., Siekierski, N., Lee, S., et al. 2018, in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 232–235
Coulton, W. R., Liu, J., McCarthy, I. G., & Osato, K. 2020, MNRAS, 495, 2531
D'Agostino, R. & Pearson, E. S. 1973, Biometrika, 60, 613
D'Agostino, R. B. 1971, Biometrika, 58, 341
de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25
Dlotko, P. 2020, in GUDHI User and Reference Manual, 3.1.1 edn. (GUDHI Editorial Board)
Dolag, K. 2015, in IAU General Assembly, Vol. 29, 2250156
Feldbrugge, J., van Engelen, M., van de Weygaert, R., Pranav, P., & Vegter, G. 2019, J. Cosmology Astropart. Phys., 2019, 052
Ferreira, T., Zhang, T., Chen, N., Dodelson, S., & LSST Dark Energy Science Collaboration. 2021, Phys. Rev. D, 103, 103535
Flaugher, B. 2005, International Journal of Modern Physics A, 20, 3121
Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
Gatti, M., Jain, B., Chang, C., et al. 2021, arXiv e-prints, arXiv:2110.10141
Gruen, D. & Brimioulle, F. 2017, MNRAS, 468, 769
Halder, A., Friedrich, O., Seitz, S., & Varga, T. N. 2021, MNRAS, 506, 2780
Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, PASJ, 72, 16
Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015, MNRAS, 453, 4384
Harnois-Déraps, J., Giblin, B., & Joachimi, B. 2019, A&A, 631, A160
Harnois-Déraps, J., Martinet, N., Castro, T., et al. 2021, MNRAS, 506, 1623
Harnois-Déraps, J., Martinet, N., & Reischke, R. 2022, MNRAS, 509, 3868
Harnois-Déraps, J., Pen, U.-L., Iliev, I. T., et al. 2013, MNRAS, 436, 540
Harnois-Déraps, J. & van Waerbeke, L. 2015, MNRAS, 450, 2857
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Heavens, A. F., Jimenez, R., & Lahav, O. 2000, MNRAS, 317, 965
Heavens, A. F., Sellentin, E., de Mijolla, D., & Vianello, A. 2017, MNRAS, 472, 4244
Hetterscheidt, M., Erben, T., Schneider, P., et al. 2005, A&A, 442, 43
Heydenreich, S., Brück, B., & Harnois-Déraps, J. 2021, A&A, 648, A74
Heymans, C., Tröster, T., Asgari, M., et al. 2021, A&A, 646, A140
Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hirschmann, M., Dolag, K., Saro, A., et al. 2014, MNRAS, 442, 2304
Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592
Ivezic, Z., Axelrod, T., Brandt, W. N., et al. 2008, Serbian Astronomical Journal, 176, 1
Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338
Jeffrey, N., Alsing, J., & Lanusse, F. 2021, MNRAS, 501, 954
Joachimi, B., Cacciato, M., Kitching, T. D., et al. 2015, Space Sci. Rev., 193, 1
Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, A&A, 638, L1
Joudaki, S., Mead, A., Blake, C., et al. 2017, MNRAS, 471, 1259
Kacprzak, T., Kirk, D., Friedrich, O., et al. 2016, MNRAS, 463, 3653
Kimura, Y. & Imai, K. 2017, Advances in Space Research, 60, 722
Kono, K. T., Takeuchi, T. T., Cooray, S., Nishizawa, A. J., & Murakami, K. 2020, arXiv e-prints, arXiv:2006.02905
Kovacev-Nikolic, V., Bubenik, P., Nikolić, D., & Heo, G. 2016, Statistical Applications in Genetics and Molecular Biology, 15, 19
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193
Lemos, P., Weaverdyck, N., Rollins, R. P., et al. 2022, arXiv e-prints, arXiv:2202.08233
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
Mandelbaum, R. 2018, ARA&A, 56, 393
Martinet, N., Castro, T., Harnois-Déraps, J., et al. 2021a, A&A, 648, A115
Martinet, N., Harnois-Déraps, J., Jullo, E., & Schneider, P. 2021b, A&A, 646, A62
Martinet, N., Schneider, P., Hildebrandt, H., et al. 2018, MNRAS, 474, 712
McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, MNRAS, 465, 2936
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. 2017 [http://arxiv.org/abs/1506.08903v7]
Oudot, S. Y. 2015, Mathematical Surveys and Monographs, Vol. 209, Persistence theory: from quiver representations to data analysis (American Mathematical Society, Providence, RI), viii+218
Parroni, C., Cardone, V. F., Maoli, R., & Scaramella, R. 2020, A&A, 633, A71
Petri, A., Liu, J., Haiman, Z., et al. 2015, Phys. Rev. D, 91, 103511
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Porqueres, N., Heavens, A., Mortlock, D., & Lavaux, G. 2021, MNRAS, 502, 3035
Porqueres, N., Heavens, A., Mortlock, D., & Lavaux, G. 2022, MNRAS, 509, 3194
Pranav, P. 2021, arXiv e-prints, arXiv:2109.08721
Pranav, P., Edelsbrunner, H., van de Weygaert, R., et al. 2017, MNRAS, 465, 4281
Pun, C. S., Xia, K., & Lee, S. X. 2018, arXiv e-prints, arXiv:1811.00252v1
Pyne, S. & Joachimi, B. 2021, MNRAS, 503, 2300
Reininghaus, J., Huber, S., Bauer, U., & Kwitt, R. 2015, in Proceedings of the IEEE conference on computer vision and pattern recognition, 4741–4748
Remus, R.-S., Dolag, K., & Hoffmann, T. 2017, Galaxies, 5, 49
Saro, A., Liu, J., Mohr, J. J., et al. 2014, MNRAS, 440, 2610
Schirmer, M., Erben, T., Hetterscheidt, M., & Schneider, P. 2007, A&A, 462, 875
Schneider, P. 1996, MNRAS, 283, 837
Schneider, P., Eifler, T., & Krause, E. 2010, A&A, 520, A116
Secco, L. F., Samuroff, S., Krause, E., et al. 2022, Phys. Rev. D, 105, 023515
Sellentin, E. & Heavens, A. F. 2016, MNRAS, 456, L132
Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P., & McCarthy, I. G. 2011, MNRAS, 417, 2020
Sheldon, E. S. & Huff, E. M. 2017, ApJ, 841, 24
Shirasaki, M. & Yoshida, N. 2014, ApJ, 786, 43
Sousbie, T. 2011, MNRAS, 414, 350
Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, arXiv e-prints, arXiv:1305.5422
Steinborn, L. K., Dolag, K., Comerford, J. M., et al. 2016, MNRAS, 458, 1013
Steinborn, L. K., Dolag, K., Hirschmann, M., Prieto, M. A., & Remus, R.-S. 2015, MNRAS, 448, 1504
Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152
Teklu, A. F., Remus, R.-S., Dolag, K., et al. 2015, ApJ, 812, 29
Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, Phys. Rev. D, 98, 043528
Uzeirbegovic, E., Geach, J. E., & Kaviraj, S. 2020, MNRAS, 498, 4021
van Daalen, M. P., Schaye, J., McCarthy, I. G., Booth, C. M., & Dalla Vecchia, C. 2014, MNRAS, 440, 2997
van de Weygaert, R., Vegter, G., Edelsbrunner, H., et al. 2013, arXiv e-prints, arXiv:1306.3640
Wasserman, L. 2018, Annual Review of Statistics and Its Application, 5, 501
Xu, X., Cisewski-Kehe, J., Green, S. B., & Nagai, D. 2019, Astronomy and Computing, 27, 34
Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45
Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., & Refregier, A. 2021, J. Cosmology Astropart. Phys., 2021, 028

Appendix A: Data compression of heatmaps

As the raw heatmaps contain 10100 entries per combination of tomographic redshift bins, a direct cosmological parameter analysis with these maps is currently impossible. We, therefore, need to explore different methods of compressing the raw data.

Our first approach to data compression is a principal component analysis (PCA). This rather simple method is highly efficient at reducing complex data to only a few manageable dimensions (see e.g. Uzeirbegovic et al. 2020). For each combination of tomographic redshift bins, we apply a PCA to the heatmap extracted from all 19 regions. We see that the PCA correctly identifies that the differences between the S/N maps of the different cosmo-SLICS are driven by the changes of the cosmological parameters Ω_m and S_8.

A PCA is still an incredibly useful tool to not only extract cosmological information from a data vector but also to understand the behaviour of the data itself better. For example, Fig. A.1 shows that the first principal component is almost exactly antiproportional to S_8, whereas the second principal component is proportional to Ω_m. Comparing these findings with Fig. A.3, we see that a high value of Ω_m leads to a large number of features in Dgm_1 (peaks) being born and dying between signal-to-noise values of 1 and 2, whereas a low value of Ω_m leads to more features in Dgm_1 being born and dying between S/N values of -0.5 and 0.5. A similar analysis for the first principal components yields the expected conclusion that a higher value of S_8 leads to more peaks being born and dying at higher S/N values and more voids being born and dying at lower S/N values. One disadvantage of PCA is the fact that it is not straightforward to include the internal covariance of the data vector: While the PCA might detect huge differences between two different cosmo-SLICS simulations, these might just be caused by the fact that this specific part of the data vector is particularly noisy, and not by differences in the cosmological signal.

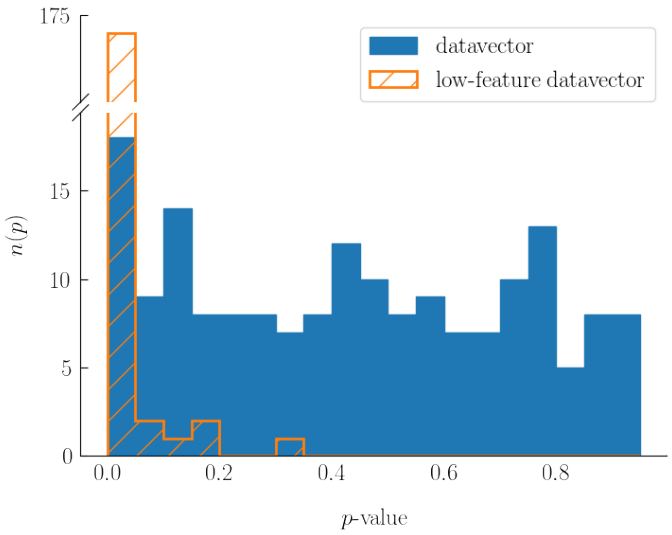


Fig. A.2. Test results for the final data-vector's Gaussianity. For each of the 180 data vector entries, we take the set of 10×124 measurements in the Covariance Training Sample and test the null hypothesis that this sample was drawn from a normal distribution using the measured skewness and kurtosis (D'Agostino 1971; D'Agostino & Pearson 1973). We then plot a histogram of the corresponding p-values (blue). If each entry of the data vector is Gaussian, then the distribution of p-values is uniform. For comparison, we show the results of the same test with a data vector that contains only points with about 10 features (orange). While the blue histogram may show small deviations from a uniform distribution (there appears to be a downward slope towards higher p-values), we believe that the assumption of a normal distribution is reasonable.

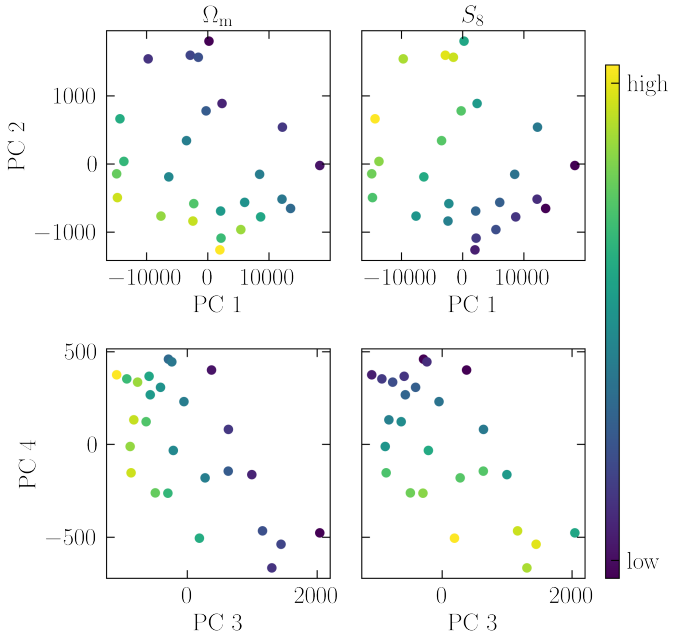


Fig. A.1. Cosmology dependence of the first four principal components. Each point in the scatter plot represents one of the 26 cosmologies of the cosmo-SLICS. In the top row, the x and y coordinates correspond to the value of the first two principal components; in the bottom row, they correspond to the values of the third and fourth principal components. In the left column, the colours represent the value of Ω_m of the respective cosmo-SLICS simulation; in the right column, they denote the value of S_8.

A more sophisticated method of data compression is the Massively Optimized Parameter Estimation and Data compression (MOPED, Heavens et al. 2000, 2017; Ferreira et al. 2021). Assuming a Gaussian likelihood, Gaussian posterior distributions and a constant covariance matrix C, this compression method preserves the entirety of the Fisher information to N_param dimensions, where N_param is the number of cosmological (and nuisance) parameters present in the inference. However, this method uses the Fisher formalism, and thus knowledge of the inverse covariance matrix C^{-1} is required. As our uncompressed data vector contains 151 500 entries and we can only estimate C with about 10^3 sets of simulations, the matrix is singular and thus not invertible (Hartlap et al. 2007). We, therefore, opted for sub-sampling our data vector and performing a MOPED compression for each individual combination of tomographic redshift bins, but doing this we neglected the information contained in the cross-correlation between different combinations of redshift bins, yielding parameter constraints that were not competitive with the ones from other data compression methods.

While MOPED is an elegant method to compress a data vector to the absolute minimum of required dimensions, this

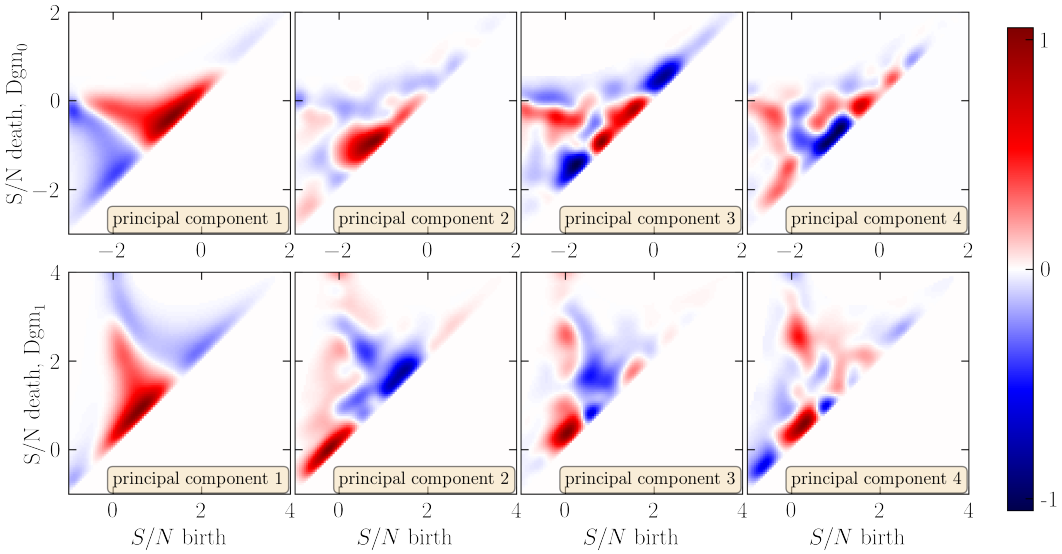


Fig. A.3. First four (normalised) principal components of the heatmaps in a principal component analysis.

also means that all additional information that was not part of this data compression gets lost. In particular, imperfect knowledge of the covariance matrix and noisy derivatives heavily affect the constraining power of MOPED. Asgari & Schneider (2015) analysed this loss of information and developed a method that is more stable with respect to changes in the covariance matrix and derivatives and offers more constraining power in the case of non-linear parameter degeneracies (like the one between Ω_m and σ_8).

All things considered, all data compression methods manage to extract a comparable amount of information out of the raw data vector (see Fig. C.1). We thus opt for the χ^2-maximiser method as the data vector obtained from this method is easiest to interpret.

Appendix B: On the observed Ω_m tension in the analysis of DES-Y1

When analysing DES-Y1 data, we observe a 3.2σ tension between the values of Ω_m estimated from 2PCF and persistent homology. In principle, there are several possible scenarios that can cause this tension. Our validation tests show that this is unlikely to be caused by a bug in the pipeline. A second possibility is that this is caused by a statistical fluctuation in the data. The third and most interesting scenario would be the presence of something unknown (and thus unaccounted for) in the DES-Y1 data. This could either be a systematic effect that we have not properly taken into account or a sign of deviations from the wCDM cosmological model that affects the topological structure of the data, but not its two-point statistics. For example, we know that persistent homology is very sensitive to primordial non-Gaussianities in the large-scale structure (Biagetti et al. 2022), which can not be detected by two-point statistics.

Appendix B.1: Investigating the severity of the tension

The fact that we achieve extraordinarily consistent results with HD+21 using a fully independent measurement and inference pipeline points to the conclusion that this tension is not caused by a bug in the pipeline. To investigate the probability of a statistical fluctuation causing this effect, we run our inference pipeline for both 2PCF and persistent homology on 100 individual lines of sight of the Covariance Training Sample. For each individual line of sight, we then estimate the tension between 2PCF and persistent homology on each cosmological parameter. The results are shown in Fig. B.1. We observe that persistent homology seems to favour higher values of Ω_m and lower values of σ_8 than 2PCF, while S_8 remains relatively unbiased. We assume that these tensions follow a normal distribution and compute its mean and variance, constructing a Gaussian fit to the values. According to this analysis, the chance that the observed bias in Ω_m is due to a statistical fluctuation is at 0.5% (2.6% for σ_8), which is still unlikely, but not as unlikely as the initial 3.2σ tension we observed suggests. Recall that no tension is observed when running our pipeline on the mock data vector constructed from all simulations of the Covariance Training Set, which is about 12× larger, suggesting that the observed tension results from statistical fluctuations that are averaged down in our validation test.

This effect certainly warrants further investigation. If something similar shows up in an analysis of KiDS-1000 (Harnois-Déraps et al. in prep., Heydenreich et al. in prep.), an investigation into potential causes for a bias in Ω_m becomes highly warranted. If, however, that analysis does not show any bias in Ω_m, we can assume that this tension is likely to be a mere statistical fluctuation in the data.

Appendix B.2: Investigating the cause of the tension

When investigating Fig. B.2, we see that an increase in Ω_m by 80% and a simultaneous decrease in S_8 by 10% almost cancels out, except for the very first point of the data vector.

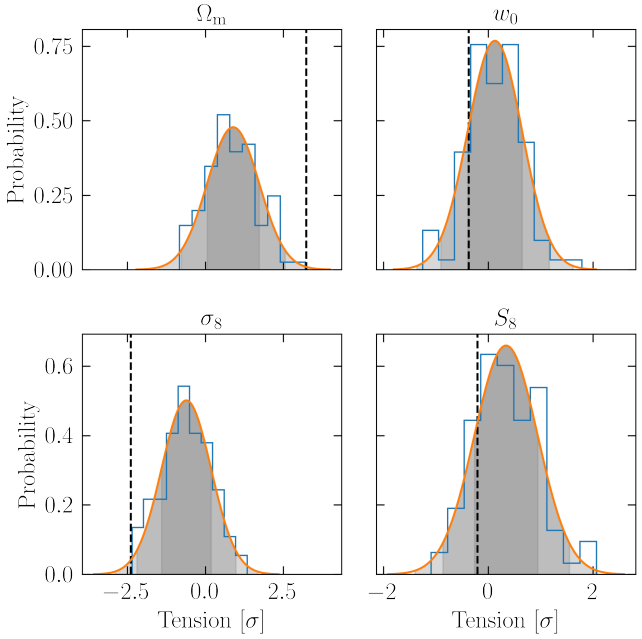


Fig. B.1. Histogram of the tensions in cosmological parameters between 2PCF and persistent homology measured on 100 individual lines of sight in the Covariance Training Set (blue) and a Gaussian fit to these values (orange). The actual tension we measured in DES-Y1 is shown by the dashed black line. No tension is observed when running our pipeline on the mock data vector constructed from all simulations of the Covariance Training Set, which is about 12× larger.

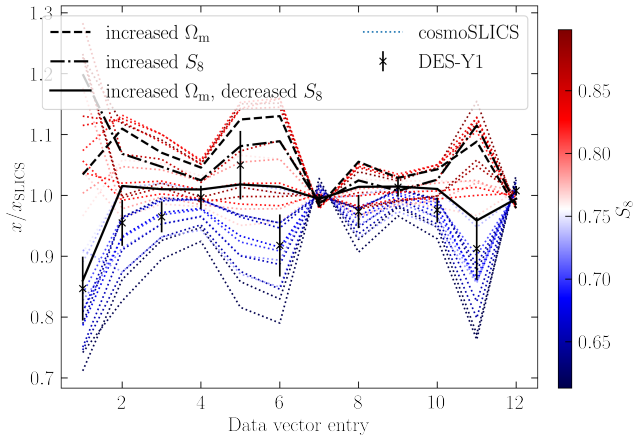


Fig. B.2. Same as Fig. 4, but we show the data vector for the individual cosmologies from our Cosmology Training Set, colour-coded by their respective value of S_8 (dotted lines) and the measured values in the DES-Y1 survey (black). For a better visibility, all values are divided by the mean of our Covariance Training Set. In addition, we predict with the GPR Emulator a data vector where Ω_m is increased by 80% with respect to the cosmology of the Covariance Training Set (dashed line), one where S_8 is increased by 10% (dash-dotted line), and one where Ω_m is increased by 80% and simultaneously S_8 is decreased by 10% (solid line).

This one corresponds to the point in the heatmap that measures the very high, very persistent peaks (see Fig. 3). We observe this behaviour consistently throughout all combinations of redshift bins. However, when we remove the first entry (corresponding to the high persistent peaks) from every tomographic bin combination, the tension between the two-point correlation function and persistent homology is reduced (see Fig. B.3) but does not vanish, and most likely, the decrease in the tension is just due to the lower constraining power. We can conclude that potentially a part of the reason for the Ω_m-tension is that we measure significantly fewer high signal-to-noise peaks in the DES-Y1 data than in the simulations, however, the tension is not caused by those peaks. The most important unmodelled systematic effect that would affect these peaks would be source-lens coupling (Martinet et al. 2018), but that one would increase the number of peaks in the simulations, not decrease it.

Although the exclusion of the high signal-to-noise peaks would reduce the Ω_m-tension without significantly affecting our constraining power on S_8, we keep our fiducial analysis choices, as switching to this analysis would constitute a major post-unblinding change.

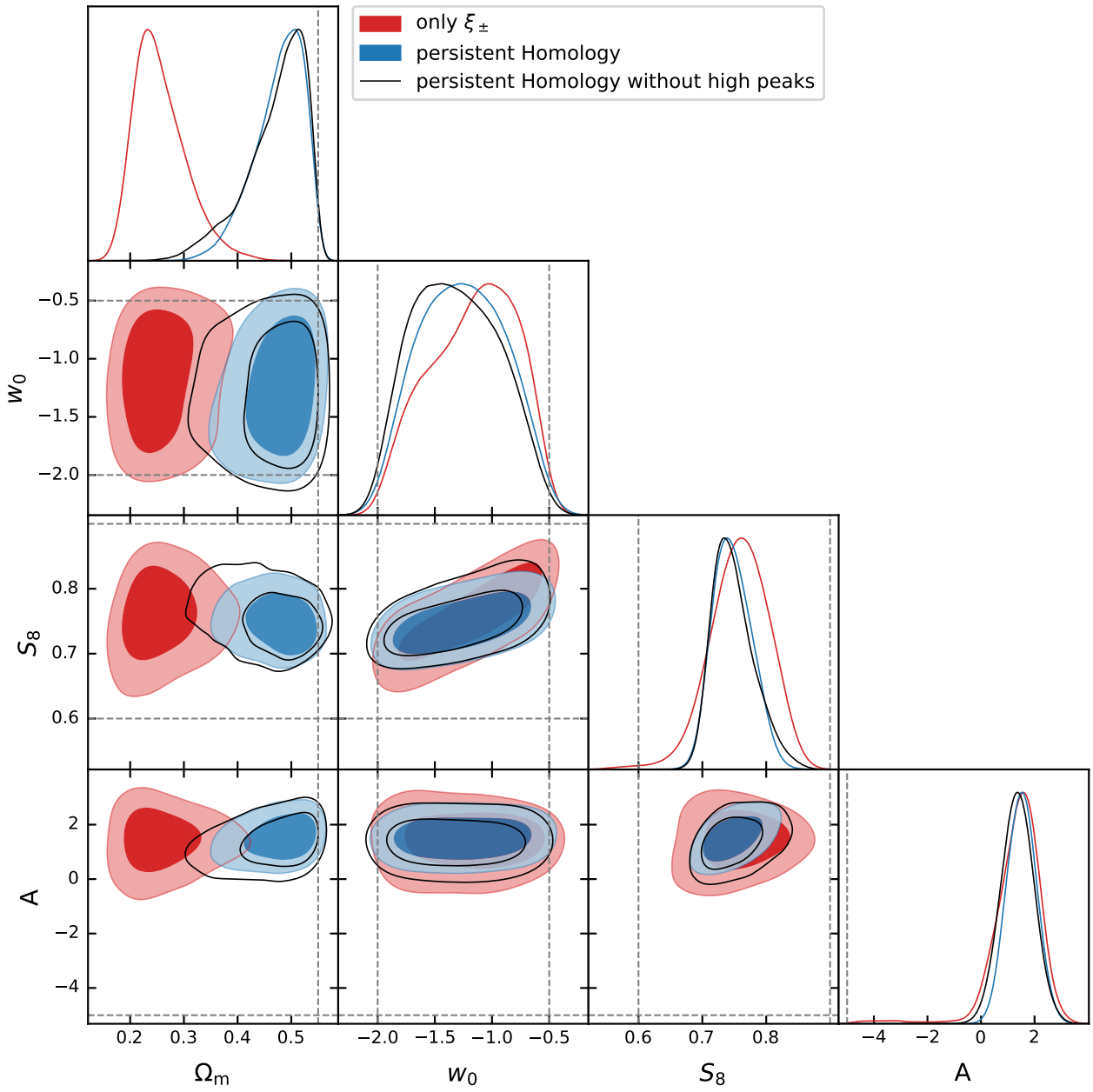


Fig. B.3. Results of our likelihood analyses for the DES-Y1 survey. We show the results for two-point statistics (red), persistent homology (blue) and report as well the constraints achieved by removing the first data point for each tomographic bin (black). The dotted lines denote the prior ranges, the black crosses denote the nodes of our Cosmology Training Set.

Appendix C: Complete parameter constraints of the MCMC

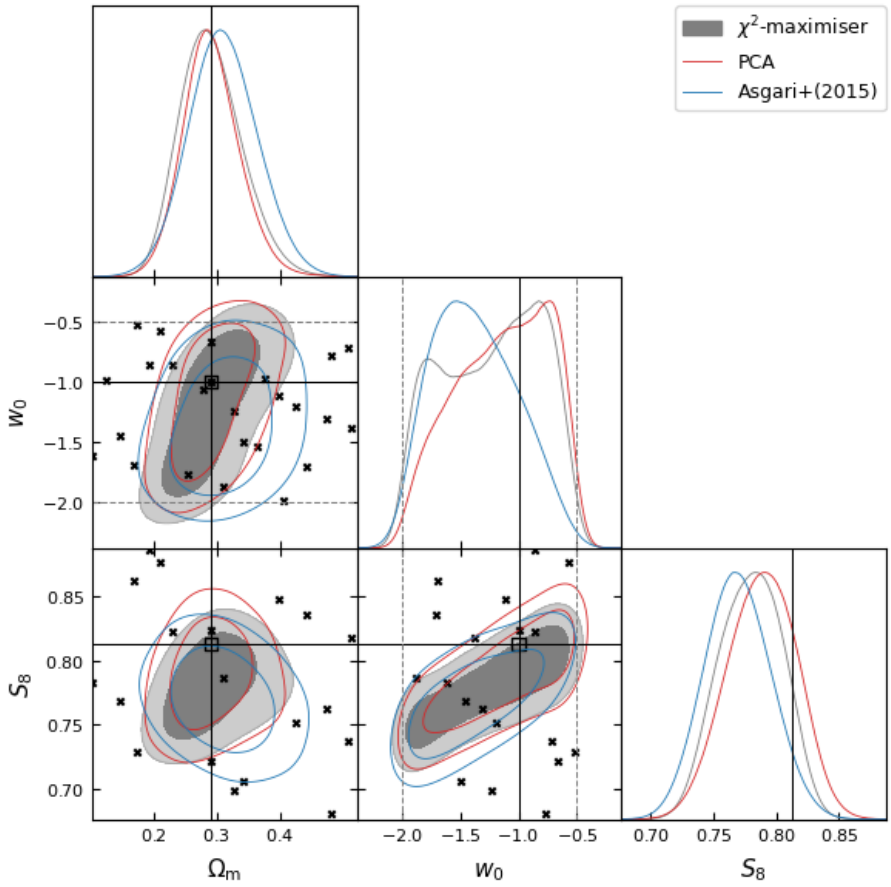


Fig. C.1. Comparison of the constraining power of different data compression methods. Our chosen method, the χ^2-maximiser is shown in grey, two alternative methods (PCA and Asgari & Schneider 2015, in red and blue, respectively).

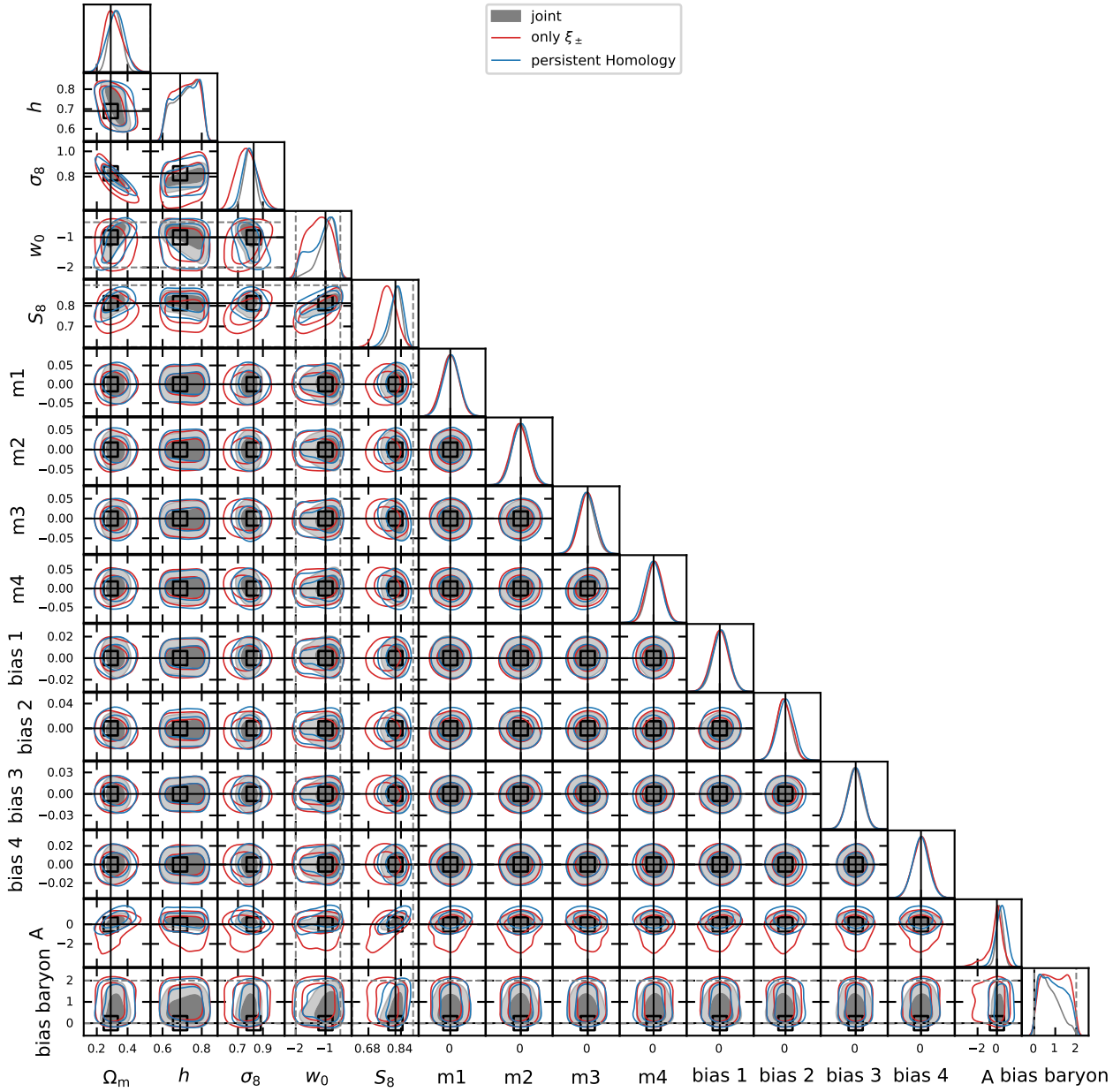


Fig. C.2. Same as Fig.10, but with all cosmological and nuisance parameters. We note that the constraints on the shear measurement bias parameters (m1 - m4) and the photometric redshift errors (bias 1 - bias 4) are dominated by the (Gaussian) prior (compare Tab. 1).

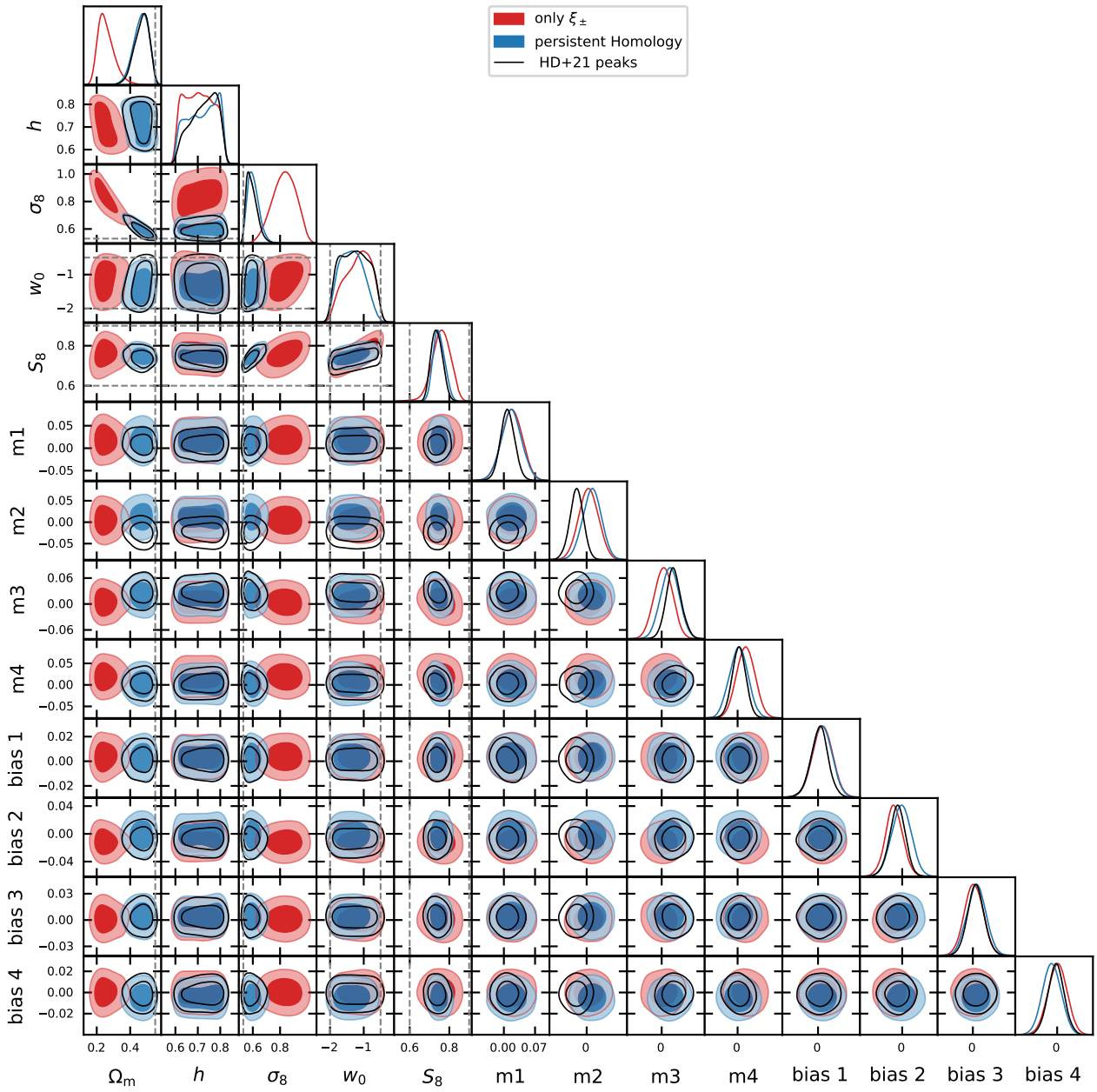


Fig. C.3. Same as Fig.11, but with all cosmological and nuisance parameters. We note that the constraints on the shear measurement bias parameters (m1 - m4) and the photometric redshift errors (bias 1 - bias 4) are dominated by the (Gaussian) prior (compare Tab. 1).

List of Figures

List of Tables

Acronyms

w**CDM** w cold dark matter model 10, 12, 102

w_0w_a**CDM** w_0w_a cold dark matter model 10

Λ**CDM** Λ cold dark matter model v, 2, 7, 10–12, 14, 23, 68, 69, 127

2pcf two-point correlation functions (Kaiser, 1992) 3, 38, 39, 46, 78, 82–86, 112, 125

3pcf three-point correlation functions ix, 39, 40, 42, 43, 87, 88, 91–93, 96–100, 102–104, 123, 124, 126, 147–149, 151, 202

BAOs baryon acoustic oscillations 2, 9, 17

CAMB Code for Anisotropies in the Microwave Background (Lewis and Challinor, 2011) 23, 89

CCD charge-coupled device 48, 49

CFHTLenS Canada-France-Hawaii Telescope Lensing Survey 4

CMB cosmic microwave background 2, 7–9, 11, 15–17, 20, 23, 27, 28, 46, 71

COSEBIs Complete Orthogonal Sets of E- and B-mode Integrals (Schneider et al., 2010) 3, 39, 45

CPU central processing unit 92

DES Dark Energy Survey (Abbott et al., 2018) viii, 4, 5, 27, 47, 65, 66, 78, 79, 81, 83, 85

DES-Y1 year-1 data of the Dark Energy Survey v, 66–68, 71, 72, 78, 80–82, 84–86, 125, 201

EoS equation of state 10–12, 65, 82, 106, 107

FFT fast Fourier Transform 72, 91, 121

FLRW Friedmann–Lemaître–Robertson–Walker vii, 8, 9, 11, 13, 34

GPR Gaussian process regression 74, 75, 77, 81–83, 86

GPU graphics processing unit 92, 97

GR general relativity 7, 8, 10, 28, 46

GRF Gaussian random fields 111, 112, 147

GUT Grand Unified Theories 16

HSC SSP Hyper Surprime-Cam Subaru Strategic Program (Aihara et al., 2018) 65, 86

HST Hubble Space Telescope 13, 49

KiDS Kilo-Degree Survey (de Jong et al., 2013) viii, 4, 5, 27, 36, 45, 47, 65, 69, 86, 126, 206

KiDS-1000 fourth data release of the KiDS (Kuijken et al., 2019) ix, 65, 69, 87, 103, 126, 151–153, 201, 202

KV450 KiDS+VIKING-450 (Wright et al., 2019) 65–68, 72, 73, 77

LISA Laser Interferometer Space Antenna 17

LSS large-scale structure v, vii, 4, 7, 17, 34, 35, 46, 68, 86

LSST Vera Rubin Observatory Legacy Survey of Space and Time (Ivezic et al., 2008) 2, 46, 65

MCMC Markov-Chain Monte Carlo viii, 63, 64, 77, 78, 88, 92, 102, 104, 106, 107

MS Millennium simulation viii, 68, 90, 96, 100, 148, 202

NFW Navarro-Frenk-White 26, 72

NLA non-linear alignment model 48

ODE ordinary differential equation 19

PCA principal component analysis 80, 102

PSF point-spread function 33, 48, 49, 65

S/N signal-to-noise 71–76, 201

SLICS Scinet light cone simulations viii, 66–68, 72–75, 77, 78, 80–86, 92, 102, 104, 121, 122, 147, 201, 202

T17 full-sky weak lensing convergence maps (Takahashi et al., 2017) viii, 69, 88, 98, 103

TATT Tidal Alignment + Tidal Torque 48

VIKING VISTA Kilo-degree Infrared Galaxy survey (Edge et al., 2013) 65

VISTA Visible and Infrared Survey Telescope for Astronomy 207

Acknowledgements

Becoming a scientist has been a naíve childhood dream of mine, which has persisted throughout the years. I am eternally grateful to everyone who helped me achieve it. I will probably not be able to thank every single person who has supported me on this journey, but I want to highlight some exceptional people.

First and foremost, I want to thank my supervisor Peter Schneider. Our first conversation was the reason that I started studying astrophysics in Bonn, and your continuous support, dedication and enthusiasm had a big impact on the last five years of my life. Thank you for believing in me. I also want to thank my second supervisor, Cristiano Porciani. Your encouragement and advice gave me the strength and confidence to power through the more frustrating phases of my research.

I also want to thank all my collaborators and co-authors, in particular Hendrik Hildebrandt, Catherine Heymans, Marika Asgari, Shahab Joudaki, Patrick Simon, Jan-Luca van den Busch, Benjamin Brück and Joachim Harnois-Déraps. Working with you has been an honour and a pleasure, and I hope to live up to the standards you set every day.

On a more personal note, I want to thank Sandra Unruh. You were the only person in the Argelander-Institute who actively encouraged me to pursue a PhD, and you have been like a mentor to me ever since. You doubled the amount of fun I had during my PhD, and you gave me a positive example of how to be both an amazing parent and a dedicated scientist. I also want to thank Jens Erler, your continuous optimism and positive attitude towards both people and life are a beacon of light in this world. Furthermore, I am very grateful to Bastian Orthen. Despite leaving the Institue too early (I really miss our after-work beers!), you have always been a good friend!

As both an inspiring example of a young scientist and a continuous source of my imposter syndrome, I want to thank Laila Linke. Your humbleness, engagement, and competence set a shining example. On the same note, I am very glad to have worked with Pierre Burger. Your tireless work to understand and tackle problems and your refusal to settle for an incomplete answer is something I genuinely admire. It has been a pleasure to grow as a scientist alongside you. I also want to thank the most recent member to our small collaboration: Lucas Porth. Thank you for all your help; your ideas have propelled me forwards many times.

I am thankful to have shared an office with Hannah Zohren. Although you gave me a bad conscience by always being there before me and yet often still leaving after me, you

made up for it with your genuine and helpful attitude. Thank you for our talks and for your leadership in the Astroclub. Just keep swimming!

My gratitude also goes out to my roommate and good friend, Davit Alkhanishvili. Thank you for keeping me sane during the continuous lockdown. I do not know many people with whom I would have enjoyed being locked in together for such a long time.

During my time at the Argelander-Institute, the work with the Astronomy on Tap team was a great source of joy for me. I want to thank all its members, in particular Sandra Unruh, Laila Linke, Toma Badescu, Maude Charmetant, Iffat Ilyas, Ivalu Christensen, Ana Mikler, Manali Jeste, Eleni Vardoulaki, Felix Pötzl, and David Ohse. On the same note, I want to thank all the members of the Astroclub for their tireless work in advertising the Master in Astronomy. Hannah Zohren, Lukas Neumann, Maude Charmetant, Evie Makroleivaditi, Cosima Eibensteiner; thank you for unforgettable times. Also, Julia Hampel, thank you so much for helping me in my exchange semester and for reviving the Astroclub.

Countless other people in the Institute have made my time here more enjoyable. Matteo Cataneo, Alexander Eggemeier, and Lucas Porth, thank you for all the coffee and for introducing me to squash. Benedetta Spina, Diana Scognamiglio, Abel Schootemeijer, Emilio Romano-Diaz, Anna Pugno, Andres Navarro, Yvonne Fichtner, Jakob van den Brok, and all of the already mentioned people, thank you for all the lunch breaks, pizza nights, and other fun events!

I also want to thank all the senior scientists, in particular Patrick Simon, Reiko Nakajima, and Emilio Romano-Diaz, as well as our IT team, Andreas Bödewig, Oliver Cordes, Ole Marggraf and Reinhold Schaaf. Thank you for supporting me and answering all kinds of stupid questions.

I further want to acknowledge all the support I have gotten from the German National Scholarship Foundation and from the International Max Planck Research School.

Ganz besonders möchte ich mich auch bei meiner Familie bedanken. Bärbel und Volker, ihr habt mich bei all meinen Vorhaben stets mit aller Kraft unterstützt, sowohl finanziell als auch emotional, und das weit über das hinaus was man von seinen Eltern normalerweise erwarten kann. Ohne euch hätte ich niemals Astrophysik studieren können, und diese ganze Arbeit wäre nicht möglich gewesen. Das werde ich euch niemals vergessen. Till und Lina, ich bin unbeschreiblich stolz auf euch. Ihr seid in vielen Bereichen des Lebens eine große Inspiration für mich und ich weiß, dass ich mich zu jeder Zeit auf euch verlassen kann.

Last but not least, I want to thank Ema Ivanuša. Thank you for always having my back, for taking my mind off things, for all our talks, for all the amazing times we had together in the last years, and for always bringing out the best in me. I am incredibly happy to have you by my side.