Université Paris-Sud

École doctorale Astronomie et Astrophysique d’Île-de-France (ED 127)
CEA Saclay, Irfu, DAp-AIM (UMR 7158)

Mémoire présenté pour l’obtention du

Diplôme d’habilitation à diriger les recherches
Discipline : Astrophysique

par

Martin KILBINGER

Cosmological parameters from weak cosmological lensing

Date de soutenance : 4 avril 2018

Lieu: CEA Saclay, Dap

Composition du jury :

| Name                      | Role                      |
|---------------------------|---------------------------|
| Alain Blanchard           | Rapporteur                |
| Martin Kunz               | Rapporteur                |
| Christophe Pichon         | Rapporteur                |
| Stéphane Plaszczynski     | Président & Examineur     |
| James G. Bartlett         | Examineur                 |
| Nicholas Kaiser           | Examineur                 |
## Contents

1 Introduction ................................. 4
   1.1 One hundred years of gravitational lensing .............................. 4
   1.2 Cosmological background .............................................. 5
      1.2.1 Standard cosmological model .................................. 5
      1.2.2 Structure formation ............................................. 6
      1.2.3 Modified gravity models ....................................... 6
   1.3 Weak cosmological lensing formalism .................................. 7
      1.3.1 Light deflection and the lens equation ........................... 7
      1.3.2 Light propagation in the universe ................................ 7
      1.3.3 Linearized lensing quantities .................................. 8
      1.3.4 Projected overdensity .......................................... 10
      1.3.5 Estimating shear from galaxies .................................. 10
      1.3.6 E- and B-modes ............................................... 11
      1.3.7 The lensing power spectrum .................................... 12
      1.3.8 Shear tomography .............................................. 13
      1.3.9 Intrinsic alignment ........................................... 14
      1.3.10 Higher-order corrections ..................................... 14

2 Shear correlation estimators .................. 15
   2.1 The shear correlation function ...................................... 15
   2.2 Derived second-order functions ..................................... 17
      2.2.1 Aperture-mass dispersion ...................................... 17
      2.2.2 Practical estimators .......................................... 18
      2.2.3 Generalisations ............................................... 18
      2.2.4 E-/B-mode mixing .............................................. 19
      2.2.5 E-/B-mode functions from a finite interval ................... 19
   2.3 Higher-order correlations ........................................... 22
      2.3.1 Third-order correlations ....................................... 22
      2.3.2 Generalized aperture-mass skewness ............................ 22
      2.3.3 Peak counts ................................................ 22

3 Inference in cosmology .......................... 25
   3.1 Covariance estimation ................................................. 25
      3.1.1 The Gaussian approximation ................................... 26
      3.1.2 Non-Gaussian contributions ................................... 26
   3.2 The likelihood function .............................................. 27
   3.3 Parameter estimation ................................................. 27
   3.4 Population Monte Carlo (PMC) ....................................... 29
      3.4.1 Convergence and effective sample size ......................... 30
      3.4.2 Performance ................................................ 30
   3.5 Model selection .................................................. 31
   3.6 Approximate Bayesian Computation (ABC) ............................ 33
1. Introduction

1.1. One hundred years of gravitational lensing

On May, 29, 1919, during a solar eclipse, the deflection of light rays of stars due to the Sun’s gravitational field was measured (Dyson et al. 1920), marking the first successful test of the theory of general relativity (GR; Einstein 1916). Only much later, in 1979 the first discovery of extra-galactic gravitational lensing was obtained, with the detection of a doubly-imaged quasar lensed by a galaxy (Walsh et al. 1979). Lensing distortions have been known since 1987 with the observation of giant arcs — strongly distorted galaxies behind massive galaxy clusters (Soucail et al. 1987). Three years later in 1990, weak gravitational lensing was detected for the first time as statistical tangential alignments of galaxies behind massive clusters (Tyson et al. 1990). It took another 10 years until, in 2000, coherent galaxy distortions were measured in blind fields, showing the existence of weak gravitational lensing by the large-scale structure, or cosmic shear (Bacon et al. 2000, Kaiser et al. 2000, Van Waerbeke et al. 2000, Wittman et al. 2000). And so, nearly 100 years after its first measurement, the technique of gravitational lensing has evolved into a powerful tool for challenging GR on cosmological scales.

All observed light from distant galaxies is subject to gravitational lensing. This is because light rays propagate through a universe that is inhomogeneous due to the ubiquitous density fluctuations at large scales. These fluctuations create a tidal gravitational field that causes light bundles to be deflected differentially. As a result, images of light-emitting galaxies that we observe are distorted. The direction and amount of distortion is directly related to the size and shape of the matter distribution projected along the line of sight. The deformation of high-redshift galaxy images in random lines of sight therefore provides a measure of the large-scale structure (LSS) properties, which consists of a network of voids, filaments, and halos. The larger the amplitude of the inhomogeneity of this cosmic web is, the larger the deformations are. This technique of cosmic shear, or weak cosmological lensing is the topic of this review.

The typical distortions of high-redshift galaxies by the cosmic web are on the order of a few percent, much smaller than the width of the intrinsic shape and size distribution. Thus, for an individual galaxy, the lensing effect is not detectable, placing cosmic shear into the regime of weak gravitational lensing. The presence of a tidal field acting as a gravitational lens results in a coherent alignment of galaxy image orientations. This alignment can be measured statistically as a correlation between galaxy shapes.

Cosmic shear is a very versatile probe of the LSS. It measures the clustering of the LSS from the highly non-linear, non-Gaussian sub-megaparsec (Mpc) regime, out to very large, linear scales of more than a hundred Mpc. By measuring galaxy shape correlations between different redshifts, the evolution of the LSS can be traced, enabling us to detect the effect of dark energy on the growth of structure. Together with the ability to measure the geometry of the Universe, cosmic shear can potentially distinguish between dark energy and modified gravity theories (Hu 1999). Since gravitational lensing is not sensitive to the dynamical state of the intervening masses, it yields a direct measure of the total matter, dark plus luminous. By adding information about the distribution of galaxies, cosmic shear can shed light on the complex relationship between galaxies and dark matter.

Since the first detection over a few square degrees of sky area a decade and a half ago, cosmic shear has matured into an important tool for cosmology. Current surveys span hundreds of square degrees, and thousands of square degrees more to be observed in the near future. Cosmic shear is a major science driver of large imaging surveys from both ground and space.

This document follows in parts my recent review “Cosmological parameters from weak cosmological lensing” (Kilbinger 2015). Various other review articles on weak gravitational lensing
1 INTRODUCTION

have covered this and related topics, see e.g. Bartelmann & Schneider (2001), Schneider et al. (2006), Hoekstra & Jain (2008), Munshi et al. (2008), Bartelmann (2010), Futamase (2015), and Mandelbaum (2017).

1.2. Cosmological background

This section provides a very brief overview of the cosmological concepts and equations relevant for weak gravitational lensing. Detailed derivations of the following equations can be found in standard cosmology textbooks, e.g. Peebles (1980), Coles & Lucchin (1996), Dodelson (2003).

1.2.1. Standard cosmological model

In the standard cosmological model, the field equations of General Relativity (GR) describe the relationship between space-time geometry and the matter-energy content of the Universe governed by gravity. A solution to these non-linear differential equations exists representing a homogeneous and isotropic universe.

To quantify gravitational lensing, however, we need to consider light propagation in an inhomogeneous universe. For a general metric that describes an expanding universe including first-order perturbations, the line element $d s^2$ is given as

$$d s^2 = \left(1 + \frac{2\Psi}{c^2}\right) c^2 d t^2 - a^2(t) \left(1 - \frac{2\Phi}{c^2}\right) d l^2,$$

(1)

where the scale factor $a$ is a function of cosmic time $t$ (we set $a$ to unity at present time $t = t_0$), and $c$ is the speed of light. The spatial part of the metric is given by the comoving coordinate $l$, which remains constant as the Universe expands. The two Bardeen gravitational potentials $\Psi$ and $\Phi$ are considered to describe weak fields, $\Psi, \Phi \ll c^2$. The potential of a lens with mass $M$ and radius $R$ can be approximated by $GM/R = (c^2/2)(R_S/R)$, where $G$ is Newton’s gravitational constant and $R_S$ is the Schwarzschild radius. The weak-field condition is fulfilled for most mass distributions, excluding only those very compact objects whose extent $R$ is comparable to their Schwarzschild radius.

In GR, and in the absence of anisotropic stress which is the case on large scales, the two potentials are equal, $\Psi = \Phi$. If the perturbations vanish, (1) reduces to the Friedmann-Lemaître-Robertson-Walker (FLRW) metric.

The spatial line element $d l^2$ can be separated into a radial and angular part, $d l^2 = d^2 + f_K^2(\chi) d \omega$. Here, $\chi$ is the comoving coordinate and $f_K$ is the comoving angular distance, the functional form of which is given for the three distinct cases of three-dimensional space with curvature $K$ as

$$f_K(\chi) = \begin{cases} 
K^{-1/2} \sin \left(\frac{K^{1/2} \chi}{2}\right) & \text{for } K > 0 \text{ (spherical)} \\
\chi & \text{for } K = 0 \text{ (flat)} \\
(-K)^{-1/2} \sinh \left[\frac{(-K)^{1/2} \chi}{2}\right] & \text{for } K < 0 \text{ (hyperbolic)}
\end{cases}$$

(2)

that are characterised by their corresponding equation-of-state relation between pressure $p$ and density $\rho$, given by the parameter $w$ as

$$p = w c^2 \rho.$$

(3)

The present-day density of each species is further scaled by the present-day critical density of the Universe $\rho_{c,0} = 3H_0^2/(8\pi G)$, for which the Universe has a flat geometry. The Hubble constant $H_0 = H(a = 1) = (\dot{a}/a)_{t = t_0} = 100 \text{ km s}^{-1}\text{Mpc}^{-1}$ denotes the present-day value of the Hubble parameter $H$, and the parameter $h \sim 0.7$ characterizes the uncertainty in our knowledge of $H_0$. The density parameter of non-relativistic matter is $\Omega_m = \rho_{m,0}/\rho_{c,0}$, which consists of cold dark matter (CDM), baryonic matter, and possibly heavy neutrinos as $\Omega_m = \Omega_c + \Omega_n + \Omega_{\nu}$. Finally, the component driving the accelerated expansion (“dark energy”) is denoted by $\Omega_{de}$. Lacking a

‡ Unless written as function of $a$, density parameters are interpreted at present time; the subscript '0' is omitted.
well-motivated physical model, the dark-energy equation-of-state parameter $w$ is often parametrized by the first or first few coefficients of a Taylor expansion, e.g. $w(a) = w_0 + w_1(1 - a)$ (Chevallier & Polarski 2001, Linder 2003). In the case of the cosmological constant, $\Omega_{de} \equiv \Omega_\Lambda$ and $w = -1$.

The sum of all density parameters defines the curvature density parameter $\Omega_K$, with $\Omega_m + \Omega_{de} + \Omega_r = 1 - \Omega_K$, where $\Omega_K = -(c/H_0)^2 K$ has opposite sign compared to the curvature $K$.

1.2.2. Structure formation

In an expanding universe, density fluctuations evolve with time. Tiny quantum fluctuations in the primordial inflationary cosmos generate small-amplitude density fluctuations. Subsequently, these fluctuations grow into the large structures we see today, in the form of clusters, filaments, and galaxy halos.

At early enough times or on large enough scales, those density fluctuations are small, and their evolution can be treated using linear perturbation theory. Once those fluctuations grow to become non-linear, other approaches to describe them are necessary — for example higher-order perturbation theory, renormalization group mechanisms, analytical models of gravitational collapse, the so-called halo model, or $N$-body simulations.

Fluctuations of the density $\rho$ around the mean density $\bar{\rho}$ are parametrized by the density contrast

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}.$$  \hspace{1cm} (4)

For non-relativistic perturbations in the matter-dominated era on scales smaller than the horizon, i.e. the light travel distance since $t = 0$, Newtonian physics suffices to describe the evolution of $\delta$ (Peebles 1980). The density contrast of an ideal fluid of zero pressure is related to the gravitational potential via the Poisson equation,

$$\nabla^2 \Phi = 4\pi G a^2 \bar{\rho} \delta.$$  \hspace{1cm} (5)

The differential equation describing the evolution of $\delta$ typically has to be solved numerically, although in special cases analytical solutions exist. The solution that increases with time is called growing mode. The time-dependent function is the linear growth factor $D_+$, which relates the density contrast at time $a$ to an earlier, initial epoch $a_i$, with $\delta(a) \propto D_+(a)\delta(a_i)$. In a matter-dominated Einstein-de-Sitter Universe, $D_+$ is proportional to the scale factor $a$. The presence of dark energy results in a suppressed growth of structures.

1.2.3. Modified gravity models

A very general, phenomenological characterisation of deviations from GR is to add parameters to the Poisson equation, and to treat the two Bardeen potentials as two independent quantities. This leads to two modified, distinct Poisson equations, which, expressed in Fourier space, are (Uzan 2006, Amendola et al. 2008)

$$k^2 \tilde{\Psi}(k, a) = 4\pi G a^2 [1 + \mu(k, a)] \rho \tilde{\delta}(k, a);$$  \hspace{1cm} (6)

$$k^2 \left[ \tilde{\Phi}(k, a) + \tilde{\Psi}(k, a) \right] = 8\pi G a^2 [1 + \Sigma(k, a)] \rho \tilde{\delta}(k, a).$$  \hspace{1cm} (7)

The tilde denotes the Fourier transform. Non-zero values of the free functions $\mu$ and $\Sigma$ represent deviations from GR. This flexible parametrization can account for a variety of modified gravity models, for example a change in the gravitational force from models with extra-dimensions as in DGP (Dvali, Gabadadze & Porrati 2000), massive gravitons (Zhytnikov & Nester 1994), $f(R)$ extensions of the Einstein-Hilbert action (de Felice & Tsujikawa 2010), or Tensor-Vector-Scalar (TeVeS) theories (Skordis 2009). Non-zero anisotropic stress is predicted from a variety of higher-order gravity theories, but also expected from models of clustered dark energy (Hu 1998, Calabrese et al. 2011). See Clifton et al. (2012) and Yoo & Watanabe (2012) for further models of modified gravity.

The above-introduced parametrization has the advantage of separating the effect of the metric on non-relativistic particles (which are influenced by density fluctuations through (6)), and light
1 INTRODUCTION

deflection (which is governed by both geometry and density fluctuations via (7), see e.g. Uzan & Bernardeau (2001), Jain & Zhang (2008)). Thus, data from galaxy clustering, redshift-space distortions, and velocity fields (testing the former relation on the one hand) and weak-lensing observations (testing the latter equation on the other hand) are complementary in their ability to constrain modified gravity models.

1.3. Weak cosmological lensing formalism

This section introduces the basic concepts of weak cosmological lensing, and discusses the relevant observables and their relationships to theoretical models of the large-scale structure. More details about those concepts can be found in e.g. Bartelmann & Schneider (2001).

1.3.1. Light deflection and the lens equation

There are multiple ways to derive the equations describing the deflection of light rays in the presence of massive bodies. An intuitive approach is the use of Fermat’s principle of minimal light travel time (Schneider et al. 1992, Schneider 1985, Blandford & Narayan 1986).

Photons propagate on null geodesics, given by a vanishing line element $ds$. In the case of GR we get the light ray travel time from the metric (1) as

$$t = \frac{1}{c} \int \left(1 - \frac{2\Phi}{c^2}\right) dr,$$

where the integral is along the light path in physical or proper coordinates $dr$. Analogous to geometrical optics, the potential acts as a medium with variable refractive index $n = 1 - \frac{2\Phi}{c^2}$ (with $\Phi < 0$), changing the direction of the light path. (This effect is what gives gravitational lensing its name.) We can apply Fermat’s principle, $\delta t = 0$, to get the Euler-Lagrange equations for the refractive index. Integrating these equations along the light path results in the deflection angle $\hat{\alpha}$ defined as the difference between the directions of emitted and received light rays,

$$\hat{\alpha} = -\frac{2}{c^2} \int \nabla_p \Phi \, dr.$$

The gradient of the potential is taken perpendicular to the light path, with respect to physical coordinates. The deflection angle is twice the classical prediction in Newtonian dynamics if photons were massive particles (von Soldner 1804).

1.3.2. Light propagation in the universe

In this section we quantify the relation between light deflection and gravitational potential on cosmological scales. To describe differential propagation of rays within an infinitesimally thin light bundle, we consider the difference between two neighbouring geodesics, which is given by the geodesic deviation equation. In a homogeneous FLRW Universe, the transverse comoving separation $x_0$ between two light rays as a function of comoving distance from the observer $\chi$ is proportional to the comoving angular distance $x_0(\chi) = f_K(\chi)\theta$ (Schneider et al. 1992, Seitz et al. 1994).

This separation vector is modified by density perturbations in the Universe. We have already seen (9) that a light ray is deflected by an amount $d\hat{\alpha} = -\frac{2}{c^2} \nabla_p \Phi(x, \chi') d\chi'$ in the presence of the potential $\Phi$ at distance $\chi'$ from the observer. Note that this equation is now expressed in a comoving frame, as well as the gradient. From the vantage point of the deflector the induced change in separation vector at source comoving distance $\chi$ is $dx = f_K(\chi - \chi') d\hat{\alpha}$ (see Fig. 1 for a sketch). The total separation is obtained by integrating over the line of sight along $\chi'$. Lensing deflections
modify the path of both light rays, and we denote with the superscript \( (0) \) the potential along the second, fiducial ray. The result is

\[
x(\chi) = f_K(\chi)\theta - \frac{2}{c^2} \int_{0}^{\chi} d\chi' f_K(\chi - \chi') \left[ \nabla_\perp \Phi(\chi', \chi') - \nabla_\perp \Phi(0)(\chi') \right].
\]

(11)

In the absence of lensing the separation vector \( x \) would be seen by the observer under an angle \( \beta = x(\chi)/f_K(\chi) \). The difference between the apparent angle \( \theta \) and \( \beta \) is the total, scaled deflection angle \( \alpha \), defining the \textit{lens equation}

\[
\beta = \theta - \alpha,
\]

(12)

with

\[
\alpha = \frac{2}{c^2} \int_{0}^{\chi} d\chi' \frac{f_K(\chi - \chi')}{f_K(\chi)} \left[ \nabla_\perp \Phi(\chi', \chi') - \nabla_\perp \Phi(0)(\chi') \right].
\]

(13)

Equation (12) is analogous to the standard lens equation in the case of a single, thin lens, in which case \( \beta \) is the source position.

1.3.3. Linearized lensing quantities  The integral equation (11) can be approximated by substituting the separation vector \( x \) in the integral by the 0th-order solution \( x_0(\chi) = f_K(\chi)\theta \) (10). This corresponds to integrating the potential gradient along the unperturbed ray, which is called the \textit{Born approximation} (see Sect. 1.3.10 for higher-order corrections). Further, we linearise the lens equation (12) and define the (inverse) amplification matrix as the Jacobian \( A = \partial \beta / \partial \theta \), which describes a linear mapping from lensed (image) coordinates \( \theta \) to unlensed (source) coordinates \( \beta \),

\[
A_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j}
\]

\[
= \delta_{ij} - \frac{2}{c^2} \int_{0}^{\chi} d\chi' \frac{f_K(\chi - \chi')}{f_K(\chi)} \frac{\partial^2 \Phi(f_K(\chi')\theta, \chi')}{\partial x_i \partial x_j}.
\]

(14)

The second term in (13) drops out since it does not depend on the angle \( \theta \).
In this approximations the deflection angle can be written as the gradient of a 2D potential, the \textit{lensing potential} $\psi$,

$$\psi(\theta, \chi) = \frac{2}{c^2} \int_0^\chi \frac{d\chi'}{f_K(\chi)} \frac{f_K(\chi - \chi')}{f_K(\chi')} \Phi(f_K(\chi') \theta, \chi').$$  \hfill (15)

With this definition, the Jacobi matrix can be expressed as

$$A_{ij} = \delta_{ij} - \partial_i \partial_j \psi,$$  \hfill (16)

where the partial derivatives are understood with respect to $\theta$. The symmetrical matrix $A$ is parametrized in terms of the scalar \textit{convergence}, $\kappa$, and the two-component spin-two \textit{shear}, $\gamma = (\gamma_1, \gamma_2)$, as

$$A = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}.$$  \hfill (17)

This defines the convergence and shear as second derivatives of the potential,

$$\kappa = \frac{1}{2} \left( \partial_1 \partial_1 + \partial_2 \partial_2 \right) \psi = \frac{1}{2} \nabla^2 \psi; \quad \gamma_1 = \frac{1}{2} \left( \partial_1 \partial_1 - \partial_2 \partial_2 \right) \psi; \quad \gamma_2 = \partial_1 \partial_2 \psi.$$  \hfill (18)

The inverse Jacobian $A^{-1}$ describes the local mapping of the source light distribution to image coordinates. The convergence, being the diagonal part of the matrix, is an isotropic increase or decrease of the observed size of a source image. Shear, the trace-free part, quantifies an anisotropic stretching, turning a circular into an elliptical light distribution.

It is mathematically convenient to write the shear as complex number, $\gamma = \gamma_1 + i\gamma_2 = |\gamma| \exp(2i\varphi)$, with $\varphi$ being the polar angle between the two shear components. Shear transforms as a spin-two quantity: a rotation about $\pi$ is the identity transformation of an ellipse (see Fig. 2 for an illustration).

In the context of cosmological lensing by large-scale structures, images are very weakly lensed, and the values of $\kappa$ and $\gamma$ are on the order of a few percent or less. Each source is mapped uniquely onto one image, there are no multiple images, and the matrix $A$ is indeed invertible.

We can factor out $(1 - \kappa)$ from $A$ (17), since this multiplier only affects the size but not the shape of the source. Cosmic shear is based on the measurement of galaxy shapes (see Sect. 4.1), and therefore the observable in question is not the shear $\gamma$ but the \textit{reduced shear},

$$g = \frac{\gamma}{1 - \kappa},$$  \hfill (19)
which has the same spin-two transformation properties as shear. Weak lensing is the regime where the effect of gravitational lensing is very small, with both the convergence and the shear much smaller than unity. Therefore, shear is a good approximation of reduced shear to linear order (see Sect. 1.3.10 for its validity).

1.3.4. Projected overdensity Since the convergence $\kappa$ is related to the lensing potential $\psi$ (15) via a 2D Poisson equation (18), it can be interpreted as a (projected) surface density. To introduce the 3D density contrast $\delta$, we apply the 2D Laplacian of the lensing potential (15) to the 3D potential $\Phi$ and add the second-order derivative along the comoving coordinate, $\partial^2 / \partial \chi^2$. This additional term vanishes, since positive and negative contributions cancel out to a good approximation when integrating along the line of sight. Next, we replace the 3D Laplacian of $\Phi$ with the over-density $\delta$ using the Poisson equation (5), and $\bar{\rho} \propto a^{-3}$. Writing the mean matter density in terms of the critical density, we get

$$\kappa(\theta, \chi) = \frac{3H_0^2 \Omega_m}{2c^2} \int_0^\chi \frac{d\chi'}{a(\chi')} f_K(\chi - \chi') f_K(\chi') \delta(f_K(\chi') \theta, \chi').$$

(20)

This expression is a projection of the density along comoving coordinates, weighted by geometrical factors involving the distances between source, deflector, and observer. In the case of a flat universe, the geometrical weight $(\chi - \chi')\chi'$ is a parabola with maximum at $\chi' = \chi/2$. Thus, structures at around half the distance to the source are most efficient to generate lensing distortions.

The mean convergence from a population of source galaxies is obtained by weighting the above expression with the galaxy probability distribution in comoving distance, $n(\chi)d\chi$,

$$\kappa(\theta, \chi) = \int_0^{\chi_{\text{lim}}} d\chi n(\chi) \kappa(\theta, \chi).$$

(21)

The integral extends out to the limiting comoving distance $\chi_{\text{lim}}$ of the galaxy sample. Inserting (20) into (21) and interchanging the integral order results in the following expression,

$$\kappa(\theta, \chi) = \frac{3H_0^2 \Omega_m}{2c^2} \int_0^{\chi_{\text{lim}}} d\chi q(\chi) f_K(\chi) \delta(f_K(\chi) \theta, \chi).$$

(22)

The lens efficiency $q$ is defined as

$$q(\chi) = \int_\chi^{\chi_{\text{lim}}} d\chi' n(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}.$$

(23)

and indicates the lensing strength at a distance $\chi$ of the combined background galaxy distribution. Thus, the convergence is a linear measure of the total matter density, projected along the line of sight with dependences on the geometry of the universe via the distance ratios, and the source galaxy distribution $n(\chi)d\chi = n(z)dz$. The latter is usually obtained using photometric redshifts (Sect. 4.5.1). We will see in Sect. 1.3.8 how to recover information in the redshift direction.

By construction, the expectation value of shear and convergence are zero, since $\langle \delta \rangle = 0$. The first non-trivial statistical measure of the distribution of $\kappa$ and $\gamma$ are second moments. Practical estimators of weak-lensing second-order statistics in real and Fourier-space are discussed in Sects. 1.3.7 and 2.1.

1.3.5. Estimating shear from galaxies In the case of cosmic shear, not the convergence but the shear is measured from the observed galaxy shapes, as discussed in this section. Theoretical predictions of the convergence (22) can be related to the observed shear using the relations (18). Further, a convergence field can be estimated by reconstruction from the observed galaxy shapes.
We can attribute an intrinsic, complex source ellipticity $\varepsilon^s$ to a galaxy. Cosmic shear modifies this ellipticity as a function of the complex reduced shear, which depends on the definition of $\varepsilon^s$. If we define this quantity for an image with elliptical isophotes, minor-to-major axis ratio $b/a$, and position angle $\phi$, as $\varepsilon = (a - b)/(a + b) \times \exp(2i\phi)$, the observed ellipticity $\varepsilon$ (for $|g| \leq 1$) is given as

$$\varepsilon = \frac{\varepsilon^s + g}{1 + g^* \varepsilon^s}.\quad (24)$$

The asterisk "$^*$" denotes complex conjugation. In the weak-lensing regime, this relation is approximated to

$$\varepsilon \approx \varepsilon^s + \gamma.\quad (25)$$

If the intrinsic ellipticity of galaxies has no preferred orientation, the expectation value of $\varepsilon^s$ vanishes, $\langle \varepsilon^s \rangle = 0$, and the observed ellipticity is an unbiased estimator of the reduced shear,

$$\langle \varepsilon \rangle = g.\quad (26)$$

This relation breaks down in the presence of intrinsic galaxy alignments (Sect. 1.3.9).

Another commonly used ellipticity estimator has been proposed by (Schneider & Seitz 1995). This estimator has a slightly simpler dependence on second moments of galaxy images, which have been widely used for shape estimation, see Sect. 4.1. However, it does not provide an unbiased estimator of $g$, but explicitly depends on the intrinsic ellipticity distribution.

In the weak-lensing regime, the shear cannot be detected from an individual galaxy. With distortions induced by the LSS of the order $\gamma \sim 0.03$, and the typical intrinsic ellipticity rms of $\sigma_\varepsilon = (\langle |\varepsilon|^2 \rangle)^{1/2} \sim 0.3$, one needs to average over a number of galaxies $N$ of at least a few hundred to obtain a signal-to-noise ratio $S/N = \gamma \times N^{1/2}/\sigma_\varepsilon$ of above unity.

### 1.3.6. E- and B-modes

The Born approximation introduced in Sect. 1.3.3 results in the definition of the convergence and shear to be functions of a single scalar potential (15). The two shear components defined in that way (18) are not independent, and the shear field cannot have an arbitrary form. We can define a vector field $\mathbf{u}$ as the gradient of the "potential" $\kappa$, $\mathbf{u} = \nabla \kappa$. By definition, the curl of this gradient vanishes, $\nabla \times \mathbf{u} = \partial_1 u_2 - \partial_2 u_1 = 0$. Inserting the relations between $\kappa$, $\gamma$ and $\psi$ (18) into this equality results in second-derivative constraints for $\gamma$. A shear field fulfilling those relations is called an *E-mode* field, analogous to the electric field. In real life however, $\mathbf{u}$ obtained from observed data is in general not a pure gradient field but has a non-vanishing curl component. The corresponding convergence field can be decomposed into its E-mode component, $\kappa^E$, and B-mode, $\kappa^B$, given by $\nabla^2 \kappa^E = \nabla \mathbf{u}$ and $\nabla^2 \kappa^B = \nabla \times \mathbf{u}$. The *B-mode* component can have various origins:

(i) Higher-order terms in the light-propagation equation (11), e.g. lens-lens coupling and integration along the perturbed light path (17) (Krause & Hirata 2010).

(ii) Other higher-order terms beyond usual approximations of relations such as between shear and reduced shear, or between shear and certain ellipticity estimators (see Sect. 4.1) (Krause & Hirata 2010).

(iii) Lens galaxy selection biases, such as size and magnitude bias (Wyithe et al. 2003, Schmidt et al. 2009a), or clustering of lensing galaxies (Bernardeau 1998, Schneider et al. 2002b).

(iv) Correlations of the intrinsic shapes of galaxies with each other, and with the structures that induce weak-lensing distortions (intrinsic alignment, Sect. 1.3.9) (Crittenden et al. 2002).

(v) Image and data analysis errors such as PSF correction residuals, systematics in the astrometry.
The astrophysical effects (i) - (iv) cause a B-mode at the percent-level compared to the E-mode. The intrinsic alignment B-mode amplitude is the least well-known since the model uncertainty is large (Capranico et al. 2013). Up to now, cosmic shear surveys do not have the statistical power to reliably detect those B-modes. Until recently, the amplitude of a B-mode detection has exclusively been used to assess the quality of the data analysis, assuming that (v) is the only measurable B-mode contributor. While this is a valid approach, it only captures those systematics that create a B-mode. A B-mode non-detection might render an observer over-confident to believe that also the E-mode is uncontaminated by systematics. Further, the ratio of B- to E-mode should not be used to judge the data quality, since this ratio is not cosmology-independent and can bias the cosmological inference of the data.

Some of my past work focused on studying and developing estimators that separate E- from B-mode in shear data. This will be presented in Sect. 2.2.

1.3.7. The lensing power spectrum  The basic second-order function of the convergence (22) is the two-point correlation function (2PCF) \( \langle \kappa(\vartheta)\kappa(\vartheta + \theta) \rangle \). The brackets denote ensemble average, which can be replaced by a spatial average over angular positions \( \vartheta \). With the assumption that the density field \( \delta \) on large scales is statistically homogeneous and isotropic, which follows from the cosmological principle, the same holds for the convergence. The 2PCF is then invariant under translation and rotation, and therefore a function of only the modulus of the separation vector between the two lines of sight \( \theta \). Expressed in Fourier space, the two-point correlation function defines the flat-sky convergence power spectrum \( P_\kappa \) with

\[
\langle \tilde{\kappa}(\ell)\tilde{\kappa}^*(\ell') \rangle = (2\pi)^2 \delta_D(\ell - \ell') P_\kappa(\ell).
\] (27)

Here, \( \delta_D \) is the Dirac delta function. The complex Fourier transform \( \tilde{\kappa} \) of the convergence is a function of the 2D wave vector \( \ell \), the Fourier-conjugate of \( \theta \). Again due to statistical homogeneity and isotropy, the power spectrum only depends on the modulus \( \ell \). For simplicity, we ignore the curvature of the sky in this expression. For lensing on very large scales, and for 3D lensing (Sect. 1.3.8), the curvature has to be accounted for by more accurate expressions (LoVerde & Afshordi 2008), or by applying spherical harmonics instead of Fourier transforms.

If the convergence field is decomposed into an E-mode \( \kappa^E \) and B-mode component \( \kappa^B \), two expressions analogous to (27) define the E- and B-mode power spectra, \( P^E_\kappa \) and \( P^B_\kappa \).

Taking the square of (22) in Fourier space, we get the power spectrum of the density contrast, \( P_\delta \), on the right-hand side of the equation. Inserting the result into (27) we obtain the flat-sky convergence power spectrum in terms of the density power spectrum as

\[
P_\kappa(\ell) = \frac{9}{4} \Omega_m^2 \left( \frac{H_0}{c} \right)^4 \int_0^{\chi_{\text{lim}}} d\chi \frac{a^2(\chi)}{a^2(\chi)} P_\delta \left( k = \frac{\ell}{f_K(\chi)}, \chi \right).
\] (28)

This simple result can be derived using a few approximations: the Limber projection is applied, which only collects modes that lie in the plane of the sky, thereby neglecting correlations along the line of sight (Limber 1953, Kaiser 1992, Simon 2007, Giannantonio et al. 2012). In addition, the small-angle approximation (expanding to first order trigonometric functions of the angle) and the flat-sky limit (replacing spherical harmonics by Fourier transforms) are used. A further assumption is the absence of galaxy clustering, therefore ignoring source-source (Schneider et al. 2002b), and source-lens (Bernardeau 1998, Hamana et al. 2002) clustering. Theoretical predictions for the power spectrum are shown in Fig. 3, using linear theory, and the non-linear fitting formulae of Takahashi et al. (2012). See Sect. 1.3.8 for the definition of the tomographic redshift bins.

The projection (28) mixes different 3D \( k \)-modes into 2D \( \ell \) wavemodes along the line-of-sight integration, thereby washing out many features present in the 3D density power spectrum. For
example, baryonic acoustic oscillations are smeared out and are not seen in the lensing spectrum (Simpson 2006, Zhang et al. 2009). This reduces the sensitivity of $P_\kappa$ with respect to cosmological parameters, for example compared to the CMB anisotropy power spectrum. Examples for some parameters are shown in Fig. 3. Then two main response modes of $P_\kappa$ for changing parameters are an amplitude change, caused by $\sigma_8$, $\Omega_m$, and $w_0$, and a tilt, generated by $n_s$ and $h$ (and, consequently, shifts are seen when varying the physical density parameters $\omega_m$ and $\omega_b$). The parameter combination that $P_\kappa$ is most sensitive to is $\sigma_8\Omega_m^\alpha$, with $\alpha \approx 0.75$ in the linear regime (Bernardeau et al. 1997).

Writing the relations between $\kappa$, $\gamma$ and the lensing potential $\psi$ (18) in Fourier space, and using complex notation for the shear, one finds for $\ell \neq 0$

$$\tilde{\gamma}(\ell) = \frac{(\ell_1 + \ell_2)^2}{\ell^2} \tilde{\kappa}(\ell) = e^{2i\beta} \tilde{\kappa}(\ell),$$

(29)

with $\beta$ being the polar angle of the wave-vector $\ell = (\ell_1, \ell_2)$, written as complex quantity. Therefore, we get the very useful fact that the power spectrum of the shear equals the one of the convergence, $P_\gamma = P_\kappa$.

The shear power spectrum can in principle be obtained directly from observed ellipticities (e.g. Hu & White 2001), or via pixellised convergence maps in Fourier space that have been reconstructed from the observed ellipticities, e.g. Seljak (1998). However, the simplest and most robust way to estimate second-order shear correlations are in real space, which we will discuss in the following section.

### 1.3.8. Shear tomography

The redshift distribution of source galaxies determines the redshift range over which the density contrast is projected onto the 2D convergence and shear. By separating source galaxies according to their redshift, we obtain lensing fields with different redshift weighting via the lens efficiency (23), thus probing different epochs in the history of the Universe with different weights. Despite the two-dimensional aspect of gravitational lensing, this allows us to recover a 3D

---

**Figure 3.** (a) The scaled tomographic convergence auto- and cross-power spectrum $\ell(\ell + 1)/(2\pi) P_{\kappa,ij}(\ell)$ for two redshift bins $(i, j) = (0, 1)$ with redshift ranges $z_0 = [0.5; 0.7]$, and $z_1 = [0.9; 1.1]$ for a Euclid-like source redshift distribution. Solid (dashed) lines correspond to the non-linear (linear) model. (b) Derivatives $d\log P_{\kappa}/d\log p_k$ of the convergence power spectrum with respect to various cosmological parameters $p_k$, as indicated in the figure. The corresponding redshift bin is $[0.9; 1.1]$. 
tomographic view of the large-scale structure. In particular, it helps us to measure subtle effects that are projected out in 2D lensing, such as the growth of structures, or a time-varying dark-energy state parameter \( w(z) \).

If we denote the redshift distribution in each of \( N_z \) bin with \( p_i, i = 1 \ldots N_z \), we obtain a new lensing efficiency \( q_i \) (23) for each case, and a resulting projected overdensity \( \kappa_i \). This leads to \( N_z(N_z - 1)/2 \) convergence power spectra \( P_{\kappa,ij} \), \( 1 \leq i \leq j \leq N_z \), including not only the auto-spectra \( (i = j) \) but also the cross-spectra \( (i \neq j) \). In (28), \( q^2 \) is replaced by the product \( q_i q_j \) (Seljak 1998, Hu 1999).

1.3.9. Intrinsic alignment Shapes of galaxies can be correlated in the absence of gravitational lensing, due to gravitational interactions between galaxies and the surrounding tidal fields. The intrinsic alignment (IA) of galaxy shapes adds an excess correlation to the cosmic shear signal that, if not taken into account properly, can bias cosmological inferences by tens of per cent. IA is difficult to account for, since it cannot simply be removed by a sophisticated galaxy selection, nor can it be easily predicted theoretically since it depends on details of galaxy formation.

Due to IA, the intrinsic ellipticity of galaxies \( \varepsilon \) no longer has a random orientation, or phase. This directly contributes to the measured two-point shear correlation function (32), as follows. The first term in (33) describes the correlation of intrinsic ellipticities of two galaxies \( i \) and \( j \). This term \( (II, \text{or shape-shape correlation}) \) is non-zero only for physically close galaxies. Its contribution to cosmic shear \( (GG, \text{or shear-shear correlation}) \), the last term in (33), can be suppressed by down-weighting or omitting entirely galaxy pairs at the same redshift (Heymans & Heavens 2003, King & Schneider 2002, King & Schneider 2003).

The second and third term in (33) correspond to the correlation between the intrinsic ellipticity of one galaxy with the shear of another galaxy. For either of these terms \( (GI, \text{or shape-shear correlation}) \) to be non-zero, the foreground galaxy ellipticity has to be correlated via IA to structures that shear a background galaxy. A lensing mass distribution causes background galaxies to be aligned tangentially. Foreground galaxies at the same redshift as the mass distribution are stretched radially towards the mass by tidal forces. Therefore the ellipticities of background and foreground galaxies tend to be orthogonal, corresponding to a negative \( GI \) correlation. For typical cosmic shear surveys with not too small redshift bins, \( GI \) dominates over \( II \). Overall, the intrinsic alignment of galaxy orientations contribute to the lensing power spectrum typically to up to 10%.

1.3.10. Higher-order corrections The approximations made in Sects. 1.3.3 and 1.3.7, resulting in the convergence power spectrum, have to be tested for their validity. Corrections to the linearised propagation equation (17) include couplings between lens structures at different redshift (lens-lens coupling), and integration along the perturbed ray (additional terms to the Born approximation). Further, higher-order correlations of the convergence take account of the reduced shear as observable. Similar terms arise from the fact that the observed size and magnitudes of lensing galaxies are correlated with the foreground convergence field (magnification and size bias; Hamana 2001, Schmidt et al. 2009b). Over the relevant scale range \( (\ell \leq 10^4) \) most of those effects are at least two orders of magnitude smaller than the first-order E-mode convergence power spectrum, and create a B-mode spectrum of similar low amplitude. The largest contribution is the reduced-shear correction, which attains nearly 10% of the shear power spectrum on arc minute scales (Bernardeau et al. 1997, Schneider et al. 1998, Dodelson et al. 2006, Krause & Hirata 2010). In Kilbinger (2010) I present simple fitting formulae that provide the reduced-shear power spectrum to 2% accuracy for \( \ell < 2 \times 10^5 \) for \( \Lambda \text{CDM} \) cosmological parameters within the WMAP7 68% error ellipsoid.

Thanks to the broad lensing kernel, the Limber approximation is very precise and deviates
Figure 4. The shear power spectrum for different approximations. Limber to first order: standard with flat-sky (L1Fl), extended for flat sky (ExtL1Fl), extended hybrid for flat sky (ExtL1FlHyb), and extended in the spherical expansion (ExtL1Sph); second-order Limber approximations: extended flat sky (ExtL2Fl), extended hybrid flat sky (ExtL2FlHyb), and extended spherical expansion (ExtL2Sph); full (exact) spherical projection (FullSph). The left panel shows the total shear power spectrum. The right panel shows the fractional difference resulting from each approximation, relative to the full spherical projection of the shear power spectrum. The two light grey curves on the top show the cosmic variance for KiDS- and Euclid-like surveys with areas of 1,500 and 15,000 square degrees, respectively. From (Kilbinger et al. 2017).

The shear power spectrum for different approximations. Limber to first order: standard with flat-sky (L1Fl), extended for flat sky (ExtL1Fl), extended hybrid for flat sky (ExtL1FlHyb), and extended in the spherical expansion (ExtL1Sph); second-order Limber approximations: extended flat sky (ExtL2Fl), extended hybrid flat sky (ExtL2FlHyb), and extended spherical expansion (ExtL2Sph); full (exact) spherical projection (FullSph). The left panel shows the total shear power spectrum. The right panel shows the fractional difference resulting from each approximation, relative to the full spherical projection of the shear power spectrum. The two light grey curves on the top show the cosmic variance for KiDS- and Euclid-like surveys with areas of 1,500 and 15,000 square degrees, respectively. From (Kilbinger et al. 2017).

from the full integration only on very large scales, for $\ell < 20$ (Giannantonio et al. 2012, Bernardeau et al. 2012). Similarly, the flat-sky approximation for the lensing power spectrum (28) and correlation function (34) provides sub-percent level accuracy on all but the very largest scales (Kilbinger et al. 2017, Kitching et al. 2017, Lemos et al. 2017). The full GR treatment of fluctuations together with dropping the small-angle approximation was also found to make a difference only on very large scales (Bernardeau et al. 2010). In Kilbinger et al. (2017) I showed that using the Limber and flat-sky approximations, current cosmological results are unaffected. I developed the second-order Limber approximation for cosmic shear, and demonstrated that this will also sufficient for future surveys, since the corresponding errors are sub-dominant compared to cosmic variance on all scales, see Fig. 4.

Many of the above mentioned corrections are more important for third-order lensing statistics (Hamana et al. 2002, Dodelson & Zhang 2005, Valageas 2014), which are presented in Sect. 2.3.1. In Fu et al. (2014) we accounted for source-lens clustering terms contributing to the lensing bispectrum. Ignoring this contamination, the parameter $\Sigma_8$ (see eq. (73)) was biased high by 0.03, which is sub-dominant compared to the statistical errors.

2. Shear correlation estimators

2.1. The shear correlation function

The most basic, non-trivial cosmic shear observable is the real-space shear two-point correlation function (2PCF), since it can be estimated by simply multiplying the ellipticities of galaxy pairs and averaging.

The two shear components of each galaxy are conveniently decomposed into tangential component, $\gamma_t$, and cross-component, $\gamma_x$. With respect to a given direction vector $\mathbf{\theta}$ whose polar
angle is $\phi$, they are defined as
\[
\gamma_t = -\Re\left( \gamma e^{-2i\phi} \right); \quad \gamma_\times = -\Im\left( \gamma e^{-2i\phi} \right).
\] (30)

The minus sign, by convention, results in a positive value of $\gamma_t$ for the tangential alignment around a mass overdensity. Radial alignment around underdensities have a negative $\gamma_t$. A positive cross-component shear is rotated by $+\pi/4$ with respect to the tangential component.

Three two-point correlators can be formed from the two shear components, $\langle \gamma_t\gamma_t \rangle$, $\langle \gamma_\times\gamma_\times \rangle$ and $\langle \gamma_t\gamma_\times \rangle$. The latter vanishes in a parity-symmetric universe, where the shear field is statistically invariant under a mirror transformation. Such a transformation leaves $\gamma_t$ invariant but changes the sign of $\gamma_\times$. The two non-zero two-point correlators are combined into the two components of the shear 2PCF (Miralda-Escude 1991),
\[
\xi_+(\theta) = \langle \gamma_t\gamma_t^* \rangle (\theta) = \langle \gamma_t\gamma_t \rangle (\theta) + \langle \gamma_\times\gamma_\times \rangle (\theta);
\xi_-(\theta) = \Re\left[ \langle \gamma_\times\gamma_\times \rangle (\theta)e^{-4i\phi} \right] = \langle \gamma_t\gamma_t \rangle (\theta) - \langle \gamma_\times\gamma_\times \rangle (\theta). \tag{31}
\]

The two components are plotted in Fig. 5. We note here that from the equality of the shear and convergence power spectrum and Parseval’s theorem, it follows that $\xi_+$ is identical to the two-point correlation function of $\kappa$.

We defined an estimator of the 2PCF in Schneider et al. (2002a) as
\[
\hat{\xi}_\pm (\theta) = \frac{\sum_{ij} w_i w_j (\varepsilon_i \varepsilon_j \pm \varepsilon_{i\times} \varepsilon_{j\times})}{\sum_{ij} w_i w_j}. \tag{32}
\]
The sum extends over pairs of galaxies $(i, j)$ at positions on the sky $\theta_i$ and $\theta_j$, respectively, whose separation $|\theta_i - \theta_j|$ lies in an angular distance bin around $\theta$. Each galaxy has a measured ellipticity $\varepsilon_i$, and an attributed weight $w_i$, which may reflect the measurement uncertainty. Using the weak-lensing relation (25) and taking the expectation value of (32), we get terms of the following type, exemplarily stated for $\xi_+$:
\[
\langle \varepsilon_i^{(s)} \varepsilon_{j\times}^{(s)*} \rangle; \langle \varepsilon_i^{(s)} \gamma_j^* \rangle; \langle \gamma_i\gamma_j^* \rangle; \text{ and } \langle \gamma_i\gamma_{j\times}^* \rangle. \tag{33}
\]
We discuss the first three terms in Sect. 1.3.9, in the context of intrinsic alignment (IA). In the absence of IA, those three terms vanish and the last term is equal to $\xi_+ (|\theta_i - \theta_j|)$. The analogous case holds for $\xi_-$. 

The main advantage of the simple estimator (32) is that it does not require the knowledge of the mask geometry, but only whether a given galaxy is within the masked area or not. For that reason, many other second-order estimators that we discuss in the following are based in this one.

The survey and mask geometry is however important to compute the covariance of (32). This influence was studied in detail in Kilbinger & Schneider (2004), where I developed a Monte-Carlo method to compute the covariance given a galaxy catalogue. This method was subsequently used for a Principal Component Analysis (Munshi & Kilbinger 2006), and Karhunen-Loève (Kilbinger & Munshi 2006) study, to examine the dependency of various survey properties on the weak-lensing information content. The same Monte Carlo method was also used in CFHTLenS to compute the covariance matrix of the 2PCF (Kilbinger et al. 2013).

Using (27) and (29), we write the 2PCF in the flat-sky approximation as Hankel transforms of the convergence power spectrum,
\[
\xi_+ (\theta) = \frac{1}{2\pi} \int d\ell \ell J_0 (\ell \theta) [P^E_\kappa (\ell) + P^B_\kappa (\ell)];
\xi_- (\theta) = \frac{1}{2\pi} \int d\ell \ell J_4 (\ell \theta) [P^E_\kappa (\ell) - P^B_\kappa (\ell)]. \tag{34}
\]
These expressions can be easily and quickly integrated numerically using fast Hankel transforms (Hamilton 2000).
The two 2PCF components mix E- and B-mode power spectra in two different ways. To separate
the two modes, a further filtering of the 2PCF is necessary, which will be discussed in the following
section.

2.2. Derived second-order functions

Apart from the 2PCF (31), other, derived second-order functions have been widely used to measure
lensing correlations in past and present cosmic shear surveys. The motivation for derived statistics
are to construct observables that (1) have high signal-to-noise for a given angular scale, (2) show
small correlations between different scales, and (3) separate into E- and B-modes. In particular the
latter property is of interest, since the B-mode can be used to assess the level of (certain) systematics
in the data as we have seen in Sect. 1.3.6.

All second-order functions can be written as filtered integrals over the convergence power
spectrum, and the corresponding filter functions define their properties.

2.2.1. Aperture-mass dispersion  Another popular statistic is the aperture-mass dispersion, denoted
as \( \langle M_{\text{ap}}^2 \rangle (\theta) \) (Fig. 6). First, one defines the aperture mass as mean tangential shear with respect to
the centre \( \theta \) of a circular region, weighted by a filter function \( Q_\theta \) with characteristic scale \( \theta \),

\[
M_{\text{ap}}(\theta, \theta') = \int d^2 \theta' Q_\theta(|\theta - \theta'|) \gamma_t(\theta') = \int d^2 \theta' U_\theta(|\theta - \theta'|) \kappa(\theta').
\]

The second equality can be derived from the relations between shear and convergence, which defines
the filter function \( U_\theta \) in terms of \( Q_\theta \) (Kaiser et al. 1994, Schneider 1996). The aperture mass is
therefore closely related to the local projected over-density, and owes its name to this fact. The
function \( U_\theta \) is compensated (i.e. the integral over its support vanishes, \( \int d^2 \theta U_\theta(\theta) = 0 \)), and filters
out a constant mass sheet \( \kappa_0 = \text{const} \), since the monopole mode (\( \ell = 0 \)) is not recoverable from
the shear (29). Two choices for the functions \( U_\theta \), and consequently \( Q_\theta \), have been widely used for
cosmic shear, a fourth-order polynomial (Schneider et al. 1998), and a Gaussian function (Crittenden
et al. 2002).

By projecting out the tangential component of the shear, \( M_{\text{ap}} \) is sensitive to the E-mode only.
One defines \( M_x \) by replacing \( \gamma_t \) with \( \gamma_x \) in (35) as a probe of the B-mode only. The variance of
(35) between different aperture centres defines the dispersion \( \langle M_{\text{ap}}^2 \rangle (\theta) \), which can be interpreted as
fluctuations of lensing strength between lines of sight, and therefore have an intuitive connection to fluctuations in the projected density contrast.

A new $N$ estimator that includes both the aperture-mass dispersion at various angular scales, and a measure of the 2PCF at one angular scale $\theta_0$, is discussed in Eifler et al. (2008). This new data vector is $\mathbf{N} = (\langle M^2_{ap}(\theta_1) \rangle, \ldots, \langle M^2_{ap}(\theta_n) \rangle, \xi_+ (\theta_0))$. It is a compromise between insensitivity to the B-mode (via the aperture-mass dispersion), and capturing the long-wavelength modes and thus maximizing the information content (through $\xi_+$). The tightest constraints on cosmological parameters are obtained with $\theta_0$ of around 10 arcmin.

### 2.2.2. Practical estimators

The aperture-mass dispersion can in principle be estimated by averaging over many aperture centres $\vartheta$. This is however not practical: The sky coverage of a galaxy survey is not contiguous, but has gaps and holes due to masking. Apertures with overlap with masked areas biases the result, and avoiding overlap results in a substantial area loss. This is particularly problematic for filter functions whose support extend beyond the scale $\theta$. One possibility is to fill in the missing data, e.g. with inpainting techniques (Pires et al. 2009b), resulting in a pixelised, contiguous convergence map on which the convolution (35) can be calculated very efficiently (Leonard et al. 2012). Alternatively, the dispersion measures can be expressed in terms of the 2PCF, and are therefore based on the estimator (32) for which the mask geometry does not play a role.

### 2.2.3. Generalisations

In fact, every second-order statistic can be expressed as integrals over the 2PCF because, as mentioned above, all are functions of $P_\kappa$, and the relation (34) can be inverted. In general, they do not contain the full information about the convergence power spectrum (Eifler et al. 2008), but separate E- and B-modes.

The general expression for an E-/B-mode separating function $X_{E,B}$ is

$$X_{E,B} = \frac{1}{2\pi} \int_0^\infty d\ell \ell P_\kappa^{E,B}(\ell) \tilde{U}^2(\ell).$$

(36)
A practical estimator using (32) is

$$\hat{X}_{E,B} = \frac{1}{2} \sum_i \vartheta_i \Delta \vartheta_i \left[ T_+ (\vartheta_i) \hat{\xi}_+ (\vartheta_i) + T_- (\vartheta_i) \hat{\xi}_- (\vartheta_i) \right].$$

(37)

Here, \(\Delta \vartheta_i\) is the bin width, which can vary with \(i\), for example in the case of logarithmic bins. The filter functions \(T_\pm\) and \(\tilde{U}^2\) are Hankel-transform pairs, given by the integral relation (Crittenden et al. 2002, Schneider et al. 2002b)

$$T_\pm(x) = \int_0^\infty dt J_{0,1}(xt)\tilde{U}^2(t).$$

(38)

This implicit relation between \(T_+\) and \(T_-\) guarantees the separation into E- and B-modes of the estimator (37).

In some cases of \(X_{E,B}\), for example for the aperture mass dispersion, the power-spectrum filter \(\tilde{U}\) is explicitly given as the Fourier transform of a real-space filter function \(U\), see e.g. (35) for the aperture mass. In other cases the functions \(T_\pm\) are constructed first, and \(\tilde{U}\) is calculated by inverting the relation (38). Model predictions of \(X_{E,B}\) can be obtained from either (36), or (37). For the latter, one inserts a theoretical model for \(\xi_\pm\), and does not need to calculate \(\tilde{U}\).

2.2.4. E-/B-mode mixing None of the derived second-order functions introduced so far provide a pure E-/B-mode separation. They suffer from a leakage between the modes, on small scales, or large scales, or both. This mode mixing comes from the incomplete information on the measured shear correlation: On very small scales, up to 10 arc seconds or so, galaxy images are blended, preventing accurate shape measurements, and thus the shape correlation on those small scales is not sampled. Large scales, at the order of degrees, are obviously only sampled up to the survey size. This leakage can be mitigated by (i) extrapolating the shear correlation to unobserved scales using a theoretical prediction (thereby potentially biasing the result), or (ii) cutting off small and/or large scales of the derived functions (thereby losing information). Figure 7 shows the example of the aperture-mass dispersion with a polynomial filter, see Kilbinger et al. (2006).

2.2.5. E-/B-mode functions from a finite interval E-/B-mode mixing can be avoided altogether by defining derived second-order statistics via suitable filter functions \(T_\pm\) (or, equivalently \(U\)). For a pure E-/B-mode separation, those filter functions need to vanish on scales where the shear correlation is missing.
The first such set of filter functions was derived geometrically, by defining the E- and B-mode of the shear field on a circle:

\[ C(\theta) = C_t(\theta) + C_\times(\theta) = \frac{1}{2\pi} \int_0^{2\pi} d\phi \left( \gamma_t + \gamma_\times \right)(\theta, \phi), \]

where the tangential and cross-components of the shear are measured with respect to the center of a circle with radius \( \theta \). By construction, \( C_t \) (\( C_\times \)) projects out the E-mode (B-mode).

If we now correlate the field \( C \) for two concentric circles with different radii \( \theta_1 < \theta_2 \), the resulting second-order E-mode (B-mode) correlations \( \langle C_t(\theta_1)C_t(\theta_2) \rangle \) (\( \langle C_\times(\theta_1)C_\times(\theta_2) \rangle \)) correlate the shear at two angular positions with minimum separation \( \theta_2 - \theta_1 \) and maximum distance \( \theta_1 + \theta_2 \). This circle statistics thus achieves E-/B-mode separation from shear correlations on a finite interval.

In practise, the shear cannot be measured on an infinitely thin line, and the circle is extended to an annulus or ring \( R \) with finite width. Two disjoint annuli are correlated to form the ring statistics (Schneider & Kilbinger 2007).

As all second-order functions, the ring statistics can be written in the forms (36) and (37). Corresponding filter functions \( U \) and \( T_\pm \) have been derived in (Schneider & Kilbinger 2007), the latter of which have finite support.

The exact form of \( T_\pm \) is given by the geometrical set-up of the two rings. This can be generalized, and \( T_\pm \) be derived detached from geometry. From the relations between \( T_+ \) and \( T_- \) and the requirement that both vanish outside a finite interval \( [\vartheta_{\text{min}}; \vartheta_{\text{max}}] \), two integral conditions are sufficient to fulfill these conditions (Schneider & Kilbinger 2007):

\[ \int_{\vartheta_{\text{min}}}^{\vartheta_{\text{max}}} d\vartheta \vartheta T_+(\vartheta) = 0 = \int_{\vartheta_{\text{min}}}^{\vartheta_{\text{max}}} d\vartheta \vartheta^3 T_+(\vartheta). \]

The corresponding relations for \( T_- \) are

\[ \int_{\vartheta_{\text{min}}}^{\vartheta_{\text{max}}} d\vartheta \vartheta T_-(\vartheta) = 0 = \int_{\vartheta_{\text{min}}}^{\vartheta_{\text{max}}} d\vartheta \vartheta^3 T_-(\vartheta). \]

The first generalised ring statistics was introduced in Eifler et al. (2010), who chose the lowest-order polynomials for \( T_+ \) to fulfill (40) (which is second order).

An optimization scheme for a general ring statistics was developed in Fu & Kilbinger (2010). In this work we wrote \( T_+ \) as linear combination of orthogonal polynomials, in this case, Chebyshev polynomials of the second kind, up to order \( N - 1 \). The two integral conditions on \( T_+ \) then become a \( (N \times 2) \) matrix equations in the expansion coefficients. It was determined that \( N = 6 \) captures most of the information of the shear correlation.

Optimisation is then performed by varying the coefficients to maximize two quantities for a given maximum angular scale \( \Psi_{\text{max}} \), the \( S/N \) of the ring statistic and the Fisher matrix figure of merit (FoM) for \( \Omega_m \) and \( \sigma_8 \). For the Fisher matrix, the (Gaussian) covariance between different scales was accounted for. Fig. 8 shows the FoM as function of \( \Psi_{\text{max}} \) for different statistics. As expected, the FoM increases with \( \Psi_{\text{max}} \) as more and more information is included, but the increase flattens out after around 20 arcmin. Compared to the original ring statistic from Schneider & Kilbinger (2007) (denoted by \( Z_+ \)), the optimised ring statistic (denoted by \( T_+ \)) achieves two to three times larger FoMs. Depending on the range of scales \( \eta = \vartheta_{\text{min}}/\vartheta_{\text{max}} \), it even outperformed the aperture-mass dispersion.

A more general question to ask is, given an interval \( [\vartheta_{\text{min}}; \vartheta_{\text{max}}] \), how can we capture all available information of the E-mode shear correlation on that interval? The information on a subset of scales, say with \( \Psi < \theta_{\text{max}} \) should be contained in the entire interval, since all information on \( [\vartheta_{\text{min}}; \vartheta_{\text{max}}] \) contains a signal with a filter function that is zero for \( \theta > \Psi \).
Such general E-/B-mode separating second-order quantities from a finite interval are the so-called COSEBIs (Complete Orthogonal Sets of E-/B-mode integrals; Schneider et al. 2010). Fig. 9 shows the COSEBIs measured with CFHTLenS. COSEBIs do not depend on a continuous angular scale parameter $\theta$, but are a discrete set of modes $E_n, B_n, n = 1, 2 \ldots$ Typically, fewer than 10 COSEBI modes are sufficient to capture all second-order E-mode information (Asgari et al. 2012).

The COSEBI modes are strongly correlated, which makes visual inspection of the data and comparison to the prediction difficult. Therefore, I compute uncorrelated data points $E_m^{\text{ortho}}$ as orthogonal transformation of the COSEBI modes $E_n, B_n = S_{mn}E_n$, where $S$ is an orthogonal matrix, $SS^T = 1$. The result is presented in the right panel of Fig. 9. Increasing modes $m$ have larger error bars, which correspond to the elements of the diagonal matrix $\Sigma$, obtained by diagonalising the COSEBIs covariance matrix $C = SS^T$.

**Figure 8.** Figure of merit (FoM) for the optimised filter function $T_+$ (blue curves), and the original ring statistic filter function $Z_+$ from SK07 (in red), as function of the maximum angular scale $\Psi_{\text{max}}$. Filled and open symbols represent $\eta \equiv \vartheta_{\text{min}}/\vartheta_{\text{max}} = 1/50$ and 1/10, respectively. The black curve with crosses corresponds to the aperture-mass with aperture diameter $\vartheta_{\text{max}}$. From Fu & Kilbinger (2010).

**Figure 9.** The first five COSEBI modes, measured in CFHTLenS, together with predictions using a WMAP7 and WMAP9 cosmology, and the CFHTLenS “Clone” $N$-body simulation. Left panel: The original COSEBI modes, from Kilbinger et al. (2013). Right panel: The corresponding tranformed, orthogonal COSEBI modes, from Fu et al. (2014).
2.3. Higher-order correlations

2.3.1. Third-order correlations  The convergence power spectrum $P_\kappa$ (28) only captures the Gaussian component of the LSS. There is however substantial complementary non-Gaussian information in the matter distribution, in particular on small scales, where the non-linear evolution of structures creates non-Gaussian weak-lensing correlations. On small and intermediate scales, these non-linear structures are the dominant contributor to non-Gaussian lensing signatures, compared to (quasi)-linear perturbations, or potential primordial non-Gaussianity. Constraints on the latter from cosmic shear alone can not compete with constraints from other probes such as CMB or galaxy clustering (Takada & Jain 2004, Pace et al. 2011, Hilbert et al. 2012).

To measure these non-Gaussian characteristics, one has to go beyond the second-order convergence power spectrum. The next-leading order statistic is the bispectrum $B_\kappa$, which is defined by the following equation:

$$\langle \tilde{\kappa}(\ell_1)\tilde{\kappa}(\ell_2)\tilde{\kappa}(\ell_3) \rangle = (2\pi)^2 \delta_D(\ell_1 + \ell_2 + \ell_3) \left[ B_\kappa(\ell_1, \ell_2) + B_\kappa(\ell_2, \ell_3) + B_\kappa(\ell_3, \ell_1) \right].$$

(42)

The bispectrum measures three-point correlations of the convergence defined on a closed triangle in Fourier space. $B_\kappa$ can be related to the density bispectrum via Limber’s equation (Cooray & Hu 2001). Other measures of non-Gaussianity are presented in Sect. 2.3.3.

The corresponding real-space weak-lensing observable is the shear three-point correlation function (3PCF) (Takada & Jain 2003, Schneider & Lombardi 2003, Zaldarriaga & Scoccimarro 2003, Benabed & Scoccimarro 2006). Correlating the two-component shear of three galaxies sitting on the vertices of a triangle, the 3PCF has $2^3 = 8$ components, and depends on three angular scales. Those eight components can be combined into four complex natural components (Schneider & Lombardi 2003, Schneider et al. 2005).

A simple estimator of the 3PCF can be constructed analogous to (32), by summing up triplets of galaxy ellipticities at binned triangles. The relations between the 3PCF and the bispectrum are complex, and it is not straightforward to efficiently evaluate those numerically. We have derived these equations in Schneider et al. (2005), and I computed them numerically in Kilbinger (2005).

2.3.2. Generalized aperture-mass skewness Most measurements and cosmological analyses of higher-order cosmic shear have been obtained using the aperture-mass skewness $\langle M_{ap}^3 \rangle$ (Pen et al. 2003, Jarvis et al. 2004, Schneider et al. 2005). $\langle M_{ap}^3 \rangle$ is the skewness of (35), and can be written as pass-band filter over the convergence bispectrum. Analogous to the second-order case, relations exist to represent $\langle M_{ap}^3 \rangle$ as integrals over the 3PCF, facilitating the estimation from galaxy data without the need to know the mask geometry. Corresponding filter functions have been found in case of the Gaussian filter (Jarvis et al. 2004).

I have contributed to define a generalization of the aperture-mass skewness. This skewness corresponds to filters with three different aperture scales, permitting to probe the bispectrum for different $\ell_1, \ell_2, \ell_3$ in Schneider et al. (2005). I have tested the increase of information from this estimator compared to the pure ”diagonal” skewness, and the aperture-mass dispersion in Kilbinger & Schneider (2005). I showed that the combination of second- and third-order statistics helps lifting parameter degeneracies, in particular the one between $\Omega_m$ and $\sigma_8$, extending earlier results (Bernardeau et al. 1997, Takada & Jain 2004) to real-space estimators.

2.3.3. Peak counts In weak-lensing data one can identify projected over-densities by isolating regions of high convergence, or enhanced tangential shear alignments. The statistics of such weak-lensing peaks are a potentially powerful probe of cosmology, since peaks are sensitive to the number of halos and therefore probe the halo mass function, which strongly depends on cosmological parameters (e.g.
A shear-selected sample of peaks is a tracer of the total mass in halos, and does not require scaling relations between mass and luminous tracers, such as optical richness, SZ or X-ray observables.

The relation between peaks and halos is complicated because of projection and noise. Several small halos in projection, or filaments along the line of sight, can produce the same lensing alignment as one larger halo. Noise in the form of intrinsic galaxy ellipticities (see Sect. 1.3.5) produces false detections, and alters the significance of real peaks (Schirmer et al. 2007). Because the number of halos strongly decreases with mass, noise typically results in an up-scatter of peak counts towards higher significance, which has to be modeled carefully.

Numerical simulations have shown a large potential of peak counts to constrain cosmological parameters (Kratochvil et al. 2010, Marian et al. 2012). Shear peaks single out the high-density regions of the LSS, and therefore probe the non-Gaussianity of the LSS. Despite peak counts being a non-linear probe of weak lensing, they require the measurement only to first order in the observed shear. Thus, this technique potentially suffers from less systematics than higher-order shear correlations. This is similar to galaxy-galaxy lensing, where shapes of background galaxies are correlated with the position of foreground objects (galaxies, but also groups and clusters). The decreased sensitivity of galaxy-galaxy and cluster lensing compared to cosmic shear has been demonstrated with CFTHLenS (Velander et al. 2014, Covone et al. 2014, Schrabback et al. 2015).

Peak counts are complementary to second-order statistics, and both probes combined are able to lift parameter degeneracies (Dietrich & Hartlap 2010, Pires et al. 2009a, Yang et al. 2011, Pires et al. 2012). In addition to peak counts, the two-point correlation function of lensing peaks carries cosmological information (Marian et al. 2013).

Theoretical predictions for peak counts are difficult to obtain, in particular at high signal-to-noise. Past approaches have been based on Gaussian random fields (Fan et al. 2010, Maturi et al. 2010). Together with my PhD student Chieh-An Lin, we introduced a new, flexible model of peak counts is based on samples of halos drawn from the mass function, which can be generated very quickly without the need to run time-consuming N-body simulations (Lin & Kilbinger 2015a).

Fig. 10 illustrates how we generate our peak count model. First, for a given redshift \( z \), halo masses \( M \) are randomly drawn from a halo mass function \( n(M,z) \), for which we choose the fitting formula from Jenkins et al. (2001). The corresponding halos are placed in a comoving volume of given field of view, where the positions \( x \) and \( y \) perpendicular to the line of sight are uniformly drawn. Halo profiles are attributed to the halos, in our case the NFW profile following Navarro et al. (1997), assuming the relation between mass and concentration from (Takada & Jain 2002). Since there is no spatial correlation between halos due to the randomization of their positions, this corresponds to the 1-halo term in the halo model, see for a review (Cooray & Sheth 2002). Next, lensing convergence and shear are computed for a given source galaxy redshift distribution, by adding up the contribution of all halos along the line of sight to a given source galaxy redshift. Shape noise is added, if desired the convergence is computed from the shear, and peaks are counted in the final \( \kappa \) map. We characterize the number of peaks with a histogram of the peak number probability function (pdf), also peak abundance, or peak function, as function of peak signal to noise ratio \( \nu \). Alternatively, we also compute the cumulative pdf (cdf) following (Dietrich & Hartlap 2010) and characterize the peak counts with the SNR at given percentiles of the cdf.

Our model makes two main assumptions: First, we claim that diffuse, unbound matter, for example in the form of filaments between halos, does not significantly contribute to the number of weak-lensing peaks. Similarly, structures such as voids are not simulated in our model. Note that models based on the halo model make the same assumptions. Second, we neglect spatial correlations between halos. Previous work has shown that correlated structure along the line of sight influences
the number of peaks by only a few percent (Marian et al. 2010). Note that in our model for a given line of sight more than one halo can contribute to the lensing signal, but in the form of random, uncorrelated halos.

We test the two hypotheses in Lin & Kilbinger (2015a) by comparing our model predictions to $N$-body simulations. First, we replace all detected halos in the simulation by analytical NFW profiles, and remove all remaining dark-matter particles. This tests our first assumption (together with the universality of the NFW profile). Next, we randomize the $x$- and $y$-positions of all halos, testing our second assumption. The result is shown in Fig. 11. Our model agrees fairly well with the $N$-body simulation, although the error bars are large due to the small field of 53.7 deg$^2$.

When replacing the $N$-body particles (blue curve) by NFW halos, the peak counts seem systematically lower (green circles). This might be due to the missing diffuse matter, or a decreased lensing strength of the NFW profiles compared to the simulated ones. For example, Lin (2016) showed that the number of peaks depends strongly on the concentration parameter.

We see a further decrease of peak counts when scrambling the halo positions (red squares), indicating the level of influence of correlated structures. Somewhat inbetween those two cases is our model, where we replace the simulated halo numbers and masses by our own draws from the mass function (cyan diamonds). Even though the number of halos in the simulation agrees well with the analytical mass function over a large range of redshift and mass (Lin & Kilbinger 2015a), the change of the number of peaks is visible. Ideally, in a cosmological analysis, one would account for the uncertainty in mass function, halo profiles, and mass-concentration relation, and marginalise over it.

The peak count model and data analysis software including parameter inference with Approximate Bayesian Computation (ABC, see Sect. 3.6) is available as the public software CAMELUS (Lin & Kilbinger 2015c).

A similar, stochastical model was proposed by Kainulainen & Marra (2009). Our model extends this earlier work by simulating an entire field of view corresponding to a real survey. This allows us to include geometrical effects such as masking, boundary effects, PSF and other spatial systematic effects.

![Figure 10. Illustration of our peak count model from creating fast simulations of uncorrelated halos. From Lin et al. (2016).](image)
residuals, and finite-field effects due to the conversion from $\kappa$ to $\gamma$. We test some of these effects in Lin et al. (2016) and Lin & Kilbinger (2018). The constraining power of peaks compared to the 2PCF was studied in Peel et al. (2017). In this paper we also demonstrated the agreement of our methods with the MICE-simulations (Fosalba et al. 2015), except for very high SNR $\nu$.

In on-going work, with PhD student Niall Jeffrey (advisor: Filipe Abdala) we examine fast approximate models such as PINOCCHIO (Taffoni et al. 2002) to create a peak count model. This could be a middle way of running simulations that take significantly less time than a N-body simulation, but does contain spatial correlations between halos to some level.

3. Inference in cosmology

Much of my work has focused on how to obtain constraints on cosmological parameters from cosmic shear observations. To that end, I studied the necessary ingredients for cosmological analyses not only for weak lensing data, but for general problems in cosmology. This includes the covariance matrix (Sect. 3.1), the likelihood function (Sect. 3.2), likelihood sampling techniques for parameter estimation (Sect. 3.3) and model comparison (Sect. 3.5), and likelihood-free inference methods (Sect. 3.6).

3.1. Covariance estimation

The covariance matrix of weak-lensing observables is an essential ingredient for cosmological analyses of cosmic shear data. Shear correlations at different scales are not independent but correlated with each other: The cosmic shear field is non-Gaussian, in particular on small scales, and different Fourier
modes become correlated from the non-linear evolution of the density field. This mode-coupling leads to an information loss compared to the Gaussian case (unless higher-order statistics are included). If not taken into account properly, error bars on cosmological parameters will be underestimated.

Additionally, even in the Gaussian case Fourier modes are spread on a range of angular scales in real space, causing shear functions to be correlated across scales. The correlation strength depends on the filter function relating the power spectrum to the real-space observable (Sects. 2.1, 2.2). The broader the filter, the stronger is the mixing of scales, and the higher is the correlation.

For an observed data vector \( \mathbf{d} = \{ d_i \}, \, i = 1 \ldots m \), the covariance matrix \( \mathbf{C} \) is defined as

\[
C_{ij} = \langle \Delta d_i \Delta d_j \rangle = \langle d_i d_j \rangle - \langle d_i \rangle \langle d_j \rangle, \tag{43}
\]

where the brackets denote ensemble average.

In a typical cosmic shear setting, the data vector \( \mathbf{d} \) consists of functions of shear correlations (e.g. the shear two-point correlation function at \( m \) angular scales \( \theta_i \), or band-estimates of the convergence power spectrum \( P_\kappa \) at \( m \) Fourier wave bands with centres \( \ell_i \)). Those functions are quadratic in the observed galaxy ellipticity \( \varepsilon \). The covariance then depends on fourth-order moments of \( \varepsilon \). From (25), one can see that the covariance can be split into three terms: The shot noise, which is proportional to \( \langle |\varepsilon|^2 \rangle^2 = \sigma^4 \), and, in the absence of intrinsic galaxy alignment (Sect. 1.3.9), only contributes to the covariance diagonal; the cosmic variance term, which depends on fourth moments of the shear; and a mixed term.

In particular the cosmic variance term is difficult to estimate since it requires the knowledge of the non-Gaussian properties of the shear field.

### 3.1.1. The Gaussian approximation

The covariance of the convergence power spectrum \( P_\kappa \) at an individual mode \( \ell \) in the Gaussian approximation is the simple expression (Kaiser 1992, Kaiser 1998, Joachimi et al. 2008)

\[
\langle (\Delta P_\kappa)^2 \rangle (\ell) = \frac{1}{f_{\text{sky}} (2\ell + 1)} \left( \frac{\sigma^2}{2\bar{n}} + P_\kappa(\ell) \right)^2. \tag{44}
\]

Here, the survey observes a fraction of sky \( f_{\text{sky}} \), with a number density of lensing galaxies \( \bar{n} \). The quadratic expression expands into shot-noise (first term), cosmic variance (second term), and a mixed term. In this Gaussian approximation, the fourth-order connected term of \( \kappa \) is zero, and the cosmic variance consists of products of terms second-order in \( \kappa \).

Analytical expressions for the Gaussian covariance of real-space second-order estimators have been obtained in (Schneider et al. 2002a, Kilbinger & Schneider 2004, Semboloni et al. 2009). The power-spectrum covariance for shear tomography is easily computed (Takada & Jain 2004).

### 3.1.2. Non-Gaussian contributions

Equation (44) can be extended to the case of a non-Gaussian convergence field, with the next-leading terms depending on the trispectrum \( T_\kappa \) (Scoccimarro et al. 1999, Takada & Jain 2004). Non-Gaussian evolution leads to a further coupling of small-scale modes with long wavelength modes that are larger than the observed survey volume. These supersurvey modes were first introduced as beat coupling in Hamilton et al. (2006), and later modeled in the halo model framework as halo sample variance (HSV; Sato et al. 2009, Kayo et al. 2013). Contrary to the other terms of the covariance that scale inversely with the survey area \( f_{\text{sky}} \), the super-survey covariance decreases faster. Therefore it is important for small survey areas (Sato et al. 2009, Takada & Hu 2013).

An alternative, non-analytic path is replacing the ensemble average in (43) by spatial averaging, and to estimate the covariance matrix from a large enough number \( n \) of independent \( N \)-body simulations. To compute the inverse of this estimator, which is needed in the likelihood function.
(see following section), the dimension of the data vector \( m \) has to be smaller than \( n \) (Anderson 2003, Hartlap et al. 2007). To reach percent-level precision for the inverse, \( n \) has be much larger than \( m \), which for future surveys with many tomographic bins means that the number of required simulations will be at least a few times \( 10^4 \) (Taylor et al. 2013, Dodelson & Schneider 2013). This was the path we chose for CFHTLenS tomography (Benjamin et al. 2013, Simpson et al. 2013, Heymans et al. 2013) and higher-order statistics (Fu et al. 2014), and COSMOS (Schrabback et al. 2010).

### 3.2. The likelihood function

To compare weak-lensing observations to theoretical predictions, one invokes a likelihood function \( L \) as the probability of the observed data \( d \) of length \( m \) given a model \( M \) with a set of parameters \( p \) of dimension \( n \).

For simplicity, in most cases, the likelihood function is modeled as an \( m \)-dimensional multi-variate Gaussian distribution,

\[
L(d|p, M) = (2\pi)^{-m/2} |C(p, M)|^{-1/2} \exp \left[ -\frac{1}{2} (d - y(p, M))^t C^{-1}(p, M) (d - y(p, M)) \right].
\]

(45)

The function \( y \) is the model prediction for the data \( d \), and depends on the model \( M \) and parameter vector \( p \). This is only an approximation to the true likelihood function, which is unknown, since shear correlations are non-linear functions of the shear field, which itself is not Gaussian, in particular on small scales.

The true likelihood function can be estimated by sampling the distribution using a suite of \( N \)-body simulations for various cosmological parameters. Because of the high computation time, this has been done only for a restricted region in parameter space (Hartlap et al. 2009, Pires et al. 2009a, Takahashi et al. 2011). For weak-lensing peak counts (Sect. 2.3.3), using our fast simulations (Lin & Kilbinger 2015a), we have sampled the true likelihood function and compared this to the Gaussian and the copula likelihood (Lin & Kilbinger 2015b). The copula is defined by transformed variables for which all one-dimensional pdfs are Gaussian. This makes the multi-variate likelihood function more Gaussian, but does not guarantee it (Sato et al. 2011).

The log-likelihood function can be approximated by a quadratic form, which is the inverse parameter covariance at the maximum point, called the Fisher matrix (Kendall & Stuart 1969, Tegmark et al. 1997). The Fisher matrix has become a standard tool to quickly assess the performance of planned surveys, or to explore the feasibility of constraining new cosmological models, e.g. (Albrecht et al. 2006). However, one has to keep in mind that the Fisher matrix is often ill-conditioned, in particular in the presence of strong parameter degeneracies, and its inversion requires a very high precision calculating of theoretical cosmological quantities, as we have shown in Wolz et al. (2012).

In most cases, the parameter-dependence of the covariance in (45) is neglected, since the computation of the covariance is very time-consuming, e.g. when derived from \( N \)-body simulations. When estimated from the data themselves, the cosmology-dependence of the covariance is missing altogether. This is a good approximation, as was shown in Eifler et al. (2009) and confirmed in Kilbinger et al. (2013), in particular when only a small region in parameter space is relevant, for example in the presence of prior information from other cosmological data.

### 3.3. Parameter estimation

Theoretical models of cosmic shear observables can depend on a large number of parameters. Apart from cosmological parameters, a number of additional, nuisance parameters might be included to characterize systematics, calibration steps, astrophysical contaminants such as intrinsic alignment,
photometric redshift uncertainties, etc. The number of such additional parameters can get very large very quickly and reach of the order a few hundred or even thousands, for example if nuisance parameters are added for each redshift bin (Bernstein 2009).

When inferring parameter constraints within the framework of a given cosmological model, one usually wants to estimate the probability of the parameter vector \( p \) given the data \( d \) and model \( M \). In a Bayesian framework, this is the posterior probability \( \pi \), which is given via Bayes’ theorem as

\[
\pi(p|d,M) = \frac{L(d|p,M)P(p|M)}{E(d|M)},
\]

which links the posterior to the likelihood function (see previous section) via the prior \( P \) and the evidence \( E \). In most cases, one wants to calculate integrals over the posterior, for example to obtain the mean parameter vector, its variance, or confidence regions. Such integrals can be written in general as

\[
I(h) = \int d^n p h(p)\pi(p|d,M),
\]

where \( h \) is a function of the parameter \( p \). To calculate the mean of the \( \alpha \)th parameter, \( I(h) = \tilde{p}_\alpha \), \( h(p) = p_\alpha \). For the variance of \( p_\alpha \), set \( h(p) = (p_\alpha - \tilde{p}_\alpha)^2 \). For a confidence region \( C \) (e.g. the 68\% region around the maximum) \( h \) is the characteristic function \( 1_C \) of the set \( C \), that is \( h(p) = 1 \) if \( p \) is in \( C \), and 0 else. Note that this does not uniquely define \( C \); there are indeed many different ways to define confidence regions.

In high dimensions, such integrals are most efficiently obtained by means of Monte-Carlo integration, in which random points are sampled from the posterior density function. Many different methods exist and have been applied in astrophysics and cosmology, such as Monte-Carlo Markov Chain (MCMC; Lewis & Bridle 2002), Population Monte Carlo (see Sect. 3.4), or multi-nested sampling (Feroz & Hobson 2008). Monte-Carlo sampling allows for very fast marginalization, for example over nuisance parameters, and projection onto lower dimensions, e.g. to produce 1D and 2D marginal posterior constraints.

MCMC provides a chain of \( N \) points \( p_j \), which under certain conditions represent a sample from the posterior distribution \( \pi \). Using this Markov chain, integrals of the form (47) can be estimated as sums over the \( N \) sample points \( p_j \),

\[
\hat{I}(h) = \frac{1}{N} \sum_{j=1}^{N} h(p_j).
\]

One caveat of this estimator is that in general, the samples are actually drawn from \( L(d|p,M)P(p|M) \), or from the unnormalised posterior. It turns out not to be trivial to estimate the normalisation (evidence) \( E \) with MCMC. However, the evidence drops out when using as MCMC method the very popular Metropolis-Hasting accept-reject algorithm (Metropolis et al. 1953, Hastings 1970). To get to the next step in the Markov Chain \( p_{i+1} \) from a previous position \( p_i \), only ratios of the posterior \( \pi(p_{i+1}|d,M)/\pi(p_i|d,M) \) are involved. Thus, (48) can be obtained without the need to compute the parameter-independent normalisation.

Other Monte-Carlo sampling techniques might provide samples under a different distribution, and (48) has to be modified accordingly, see for example the following section with the case of PMC.

Alternatively, in a frequentist framework, one can minimize the function \( \chi^2 = -2\ln L \). Marginalisation can be performed with the so-called profile likelihood method (Planck Collaboration et al. 2014). Frequentist minimisation is equivalent to Bayesian inference with flat priors on all parameters.

Cosmic shear using current data is sensitive to only a few cosmological parameters, in particular \( \Omega_m \) and \( \sigma_8 \). Shear tomography is beginning to obtain interesting results on other parameters such
as $\Omega_K$, or $w$. For parameters that are not well constrained by the data, for example $\Omega_b$ or $h$, the (marginal) posterior is basically given by the prior density. Therefore, the prior should be chosen wide enough to not restrict other parameters, and to not result in overly optimistic constraints.

### 3.4. Population Monte Carlo (PMC)

In Wraith et al. (2009) and Kilbinger et al. (2010) we develop for cosmology Population Monte Carlo, a sampling technique based on iterative importance sampling (Cappé et al. 2008, Cappé et al. 2004). Importance sampling provides samples under a posterior distribution $\pi$, but samples from a different distribution, the so-called proposal or importance function $q$, choosen to be a simple function from where samples can be generated easily.

To provide samples under $\pi$, we re-write (47) as

$$I(h) = \int d^n p h(p) \pi(p|d,M) = \int d^n p h(p) \frac{\pi(p|d,M)}{q(p)} q(p). \quad (49)$$

This identity holds for any $q$ whose support includes the support of $\pi$, and for functions $h$ whose expectation $I(h)$ is finite.

This expression is estimated by sampling under the proposal $q$. The Monte-Carlo estimator (48) is simply modified to account for the additional term in the integral,

$$\hat{I}(h) = \frac{1}{N} \sum_{j=1}^{N} h(p_j) w_j; \quad w_j = \frac{\pi(p_j|d,M)}{q(p_j)}. \quad (50)$$

The $w_j$ are called importance importance weights.

To estimate the unnormalised posterior (50), can be modified to include the self-normalised importance ratios

$$\hat{I}(h) = \frac{1}{N} \sum_{j=1}^{N} h(p_j) \bar{w}_j; \quad \bar{w}_j = \frac{w_j}{\sum_{i=1}^{N} \bar{w}_i}, \quad (51)$$

where the $\bar{w}_j$ are the normalised importance weights. This circumvents the necessity to compute the normalisation $E$. Contrary to the case of MCMC however, PMC does provide a robust estimate of the evidence that comes at no extra cost, as I will show below.

The performance of importance sampling depends strongly on the choice of $q$. If the importance function does not well match $\pi$, many sampled points receive a very low weight, leading to a very poor efficiency and large variance of the estimator. PMC proposes to solve this problem by iteratively adapting the importance function to the posterior: A sequence of importance functions $q^t$, $t = 1 \ldots T$ aims to approximate the target $\pi$.

The approximation is quantified by the Kullback-Leibler divergence, or relative entropy $K$,

$$K(\pi||q^t) = \int d^n p \log \left( \frac{\pi(p|d,M)}{q^t(p)} \right) \pi(p), \quad (52)$$

which is an asymmetric measure of the similarity between two distributions. If $K$ is 0, the two distributions are identical almost everywhere, whereas for $K \to 1$, $\pi$ and $q$ are very different.

The PMC algorithm adjusts the density $q^t$ incrementally such that the divergence decreases with progressive iterations, or $K(\pi||q^{t+1}) < K(\pi||q^t)$.

Following Cappé et al. (2008), we use a variant of the expectation-maximization (EM) algorithm to obtain updated parameters that determine the proposal function, for which we choose a mixture (weighted sum) of Gaussian or Student-t distributions. The parameters are mixture weights, mean, and covariance matrix for the Gaussian or Student-t components of the mixture model. Cappé et al. (2008) derived closed-form solutions for these parameters, which are given as integrals over
the posterior \( \pi \). In iteration \( t + 1 \), these integrals are approximated by evaluating their importance sampling estimator using the samples and importance weights from the previous iteration \( t \). The details of the algorithm are summarized in Wraith et al. (2009).

### 3.4.1. Convergence and effective sample size

Formally, there is no convergence criterium for PMC. Eq. (51) is an unbiased estimator of \( I(h) \) if the support of the proposal \( q \) covers \( \pi \). Unlike the case of MCMC, there is no burn-in phase and no asymptotic convergence of the chain towards the posterior distribution. However, for a badly adapted proposal, the estimate might be very noisy. Therefore, a criterium is introduced that monitors the improvement of the adaption. An estimate of \( \exp(-K) \) is the so-called normalised perplexity for iteration \( t \), \( p^t \),

\[
p^t = \exp(H^t / N),
\]

where \( H^t \) is the Shannon entropy of the normalised weights,

\[
H^t = -\sum_{j=1}^{N} \bar{w}^t_j \log \bar{w}^t_j.
\]

The perplexity ranges between 0 and 1. Values of \( p \) close to unity indicate good agreement between the importance function and the posterior.

A stopping criterium for a PMC run can be defined to be the iteration for which \( p \) exceeds a pre-determined threshold. Typically, a final importance run with a higher number of sample points than for each previous iteration is being carried out using the proposal from the last iteration. This will be the final sample use for inference, to estimate (51).

A quantity related to the efficiency of PMC is the effective sample size, \( ESS^t \),

\[
ESS^t = \left( \sum_{j=1}^{N} \left\{ \bar{w}^t_j \right\}^2 \right)^{-1},
\]

where \( 1 \leq ESS^t \leq N \). The effective sample size can be interpreted as the number of sample points with non-zero weight. The ESS can directly be compared with the number of sampled points of a Markov Chain multiplied with the acceptance rate.

### 3.4.2. Performance

Fig. 12 shows perplexity and effective sample size as number of sample points, for a PMC run using CMB anisotropy data from WMAP5 (Hinshaw et al. 2009). We sample seven cosmological parameters of a \( w \)CDM model, see Wraith et al. (2009) for details. PMC is run for 10 iterations, using a mixture Gaussian importance function. Each point on the figure shows the value after the corresponding iteration. Each of the first 9 iterations is performed with 10,000 sample points, except for the final one, which has a number five times larger, to reduce Monte-Carlo noise.

After five iterations, or 50,000 sampled points, the perplexity reaches values of 0.9 and higher. The normalised ESS increases up to the last iteration and exceeds 0.6. This is much higher than the typical MCMC acceptance rate of 0.25, even when taking into account the number of samples from all iterations (130,000 in total).

The number of samples needed for PMC is of the same order of magnitude as for MCMC. The total computing time is therefore similar. However, importance sampling separates the (typically time-consuming) evaluation of the posterior, or the likelihood function, from the sampling. Obtaining the sample points (= draws from Gaussian or Student-t distributions) for each iteration is very fast. The computation of the importance weights, which involves evaluation of the likelihood at the sampled points, can then be performed in parallel. For a number \( n \) of CPUs, the wall-clock time gain is nearly \( n \).
Figure 12. Perplexity (left panel; eq. (53)) and effective sample size ESS divided by the number of sample points, ESS/N (right panel; eq. (55)), as a function of the cumulative sample size $N$. The likelihood is WMAP5 for a flat ΛCDM model with six parameters. From Wraith et al. (2009).

In Wraith et al. (2009) we show for a toy example in 10 dimension that PMC is capable to well sample the tails of a narrow distribution. The variance of the estimator (51) for various functions $h$ such as the mean and credible regions is typically smaller than for MCMC.

The implementation of Population Monte-Carlo for cosmology is available as the public software CosmoPMC (Kilbinger et al. 2012).

3.5. Model selection

The previous section discussed estimating parameters within a given cosmological model, with the goal to measure mean and error bars of model parameters. Taking one step back, one can ask the more fundamental question what the best model is that describes the observations. This is a qualitative different step and needs to compare different models, which is independent of estimating parameters of those models. Such a comparison has to account for the ability of models to describe the data, and the model complexity. This is achieved naturally with posterior probabilities from a Bayesian analysis.

The Bayesian evidence $E$ is the denominator in Bayes’ theorem (46). Since the posterior, being a probability distribution function, is normalised, this normalization is identical to $E$. Re-writing (46) and integrating yields the evidence as the integral over the unnormalised posterior (likelihood $\times$ prior)

$$E(d|M) \int d^n p \pi(p|d, M) = E(d|M) = \int d^n p L(d|p, M)P(p|M).$$

(56)

This integral over the entire parameter space can be interpreted as a measure of the overall model probability $M$ given the data $d$.

The likelihood $L$ accounts for the goodness of the fit with respect to the data, quantifying the data fidelity of the model. The better the agreement of the data with the model, the higher the likelihood and thus the larger $E$ will be.

The Bayesian evidence also crucially depends on the prior $P$. The larger the parameter space, the smaller the amplitude of $P$ will be in general, since $P$ is a normalised probability distribution, and thus the smaller the evidence becomes. This penalizes models that have a large parameter space,
Jeffrey’s scale to quantify the ‘strength of evidence’ for a corresponding range of the Bayes factor $B_{12}$ in (57), assuming $B_{12} > 1$.

| $\ln B_{12}$ | $B_{12}$ | Strength |
|--------------|----------|----------|
| $< 1$        | $< 2.7$  | Inconclusive |
| $1 \ldots 2.5$ | $2.7 \ldots 12$ | Weak |
| $2.5 \ldots 5$ | $12 \ldots 150$ | Moderate |
| $> 5$        | $> 150$  | Strong |

and that are thus not very predictive: The predictability of a model $M$ given some data $d$ only makes sense when compared to a prior knowledge. A model has low predictability if it requires fine-tuning of parameters, i.e. when the posterior is very concentrated compared to the prior.

A good model is one for which the prior distribution of parameters closely matches “reality”, that is the posterior distribution obtained from a measurement $d$. This is true in particular when the prior is obtained from first principles or physical knowledge of the system in question. To use the Bayesian evidence as model probability, this needs to be the case, since the prior is an integral part of the model, and thus has to be physically well-motivated instead of chosen in an ad-hoc way (Efstathiou 2008). For arbitrary priors, the results of Bayesian model comparison are arbitrary.

The Bayesian evidence has thus a built-in Occam’s razor term, penalizing models with high complexity (in the form of having many additional parameters, or a large, “wasted” parameter space), see also Berger & Jeffreys (1992) and MacKay (2002).

To compare two models $M_1$ and $M_2$, we use the Bayes factor (Jeffreys 1939),

$$B_{12} = \frac{E_1}{E_2} := \frac{E(d|M_1)}{E(d|M_2)}.$$  \hfill (57)

If $B_{12}$ is larger (smaller) than unity, the data favour model $M_1$ ($M_2$) over the alternative model. To quantify the “strength of evidence” contained in the data, Jeffreys (1961) introduced an empirical scale, see Table 1. This is only a rough guideline for decision making and of course the proposed boundaries are mostly subjective, albeit connected with information theory. I refer the reader to Trotta (2008) for a pedagogical review about the Bayesian evidence and its use for model comparison.

The integral (56) is readily approximated with the importance sampling estimator (50), where the function $h$ is unity,

$$\hat{E}(d, M) = \frac{1}{N} \sum_{j=1}^{N} w_j.$$  \hfill (58)

Note that here the unnormalised importance weights are used.

Below I show an example from Kilbinger et al. (2010), comparing inflationary models of the power spectrum of primordial perturbations. I use data from CMB anisotropies (Hinshaw et al. 2009), SNIa (Kowalski et al. 2008) and BAO (Eisenstein et al. 2005).

I model the primordial scalar perturbations power spectrum as

$$P_\delta(k) \propto k^{n_s} e^{\frac{1}{2} \alpha_s \ln(k/k_0)},$$  \hfill (59)

with the parameters $n_s$ being the scalar spectral index, and $\alpha_s$ the “running” of the index. The pivot scale $k_0$ is fixed to $k_0 = 0.002$ Mpc$^{-1}$. In addition, tensor modes (gravitational waves) have the power spectrum

$$P_t(k) \propto k^{n_t},$$  \hfill (60)
Table 2. Prior ranges for primordial model comparison. The prior ranges for primordial parameters are derived from the slow-roll approximation.

| Parameter | Description                | Min. | Max. |
|-----------|----------------------------|------|------|
| $n_s$     | Scalar spectral index      | 0.39 | 1.2  |
| $\alpha_s$| Running of spectral index  | -0.2 | 0.033|
| $r$ (lin. prior) | Tensor-to-scalar ratio | 0    | 1.65 |
| $\ln r$ (log. prior) | Tensor-to-scalar ratio | -80  | 0.50 |

with tensor spectral index $n_t$. The ratio between tensor and scalar perturbation spectra at scale $k_0$ is denoted by $r$.

The standard $\Lambda$CDM model $M_2$ has $\alpha_s = n_t = r = 0$, with only $n_s$ being a free parameter. Against $M_2$ we test various models $M_1$, where we keep combinations of $\alpha_s$ and $r$ free. The tensor index $n_t$ is unconstrained by current data and therefore not included. In addition, we test the Harrison-Zel’dovich model of a scale-free power spectrum with $n_s$. This model has meanwhile been ruled out at the 5σ level by Planck (Planck Collaboration et al. 2016) for the standard $\Lambda$CDM model, but this significance decreases for extended models.

To obtain physically motivated priors on the power-spectrum parameter, we consider the slow-roll approximation of inflation. This approximation provides an infinite hierarchy of flow equations describing the dynamics of the single scalar field which drives inflation (Peiris & Easther 2006). The slow-roll parameters to first order are $\epsilon$ and $\eta$, which are given in terms of the inflaton potential and Hubble parameter during inflation.

Slow-roll conditions are satisfied when $0 \leq \epsilon \leq 0.1$ and $|\eta| \leq 0.1$. Although the numerical values are approximate, they are natural limits for the validity of the Taylor-expansion of the power spectrum $P(k)$ in $\ln(k/k_0)$, see Martin & Ringeval (2006). From these limits, and their relations to the power-spectrum parameters we derive priors on the latter, see Table 2. We choose uninformative (flat) priors for simplicity. With that we have the ingredients for a meaningful Bayesian model comparison analysis, which are well-defined models with physically motivated priors.

In Fig. 13 we show the Bayes factor of various models $M_1$ with respect to the standard, reference model $M_2$ (the flat $\Lambda$CDM universe with $n_s = \text{const}$), as function of the number of parameters $n_{\text{par}}$ for each model $M_1$. A running spectral index is favoured weakly, all other cases are disfavoured. The evidence against the Harrison-Zel’dovich model ($n_s = 1$) is only weak, even though the measured value of $n_s = 0.9622 \pm 0.0145$ is inconsistent with unity at the 2.6σ level for this data set. This shows the importance of Bayesian model comparison when compared to the significance of marginalised error bars using a single model, when making statements about ruling out models.

Tensor perturbations are moderately disfavoured. However, as example of the influence of the prior choice, we use in addition a model with a flat prior in $\ln r$ instead of $r$. As lower limit we choose $-80$, corresponding to the energy scale of Big Bang Nucleosynthesis as a conservative lower limit of the inflation energy scale (Parkinson et al. 2006). The large prior of the logarithmic tensor-to-scalar ratio causes this model to be strongly disfavoured.

3.6. Approximate Bayesian Computation (ABC)

Approximate Bayesian Computation (Marin et al. 2011, Cameron & Pettitt 2012) is a so-called likelihood-free statistical inference method. That means the evaluation of a likelihood function is bypassed to obtain samples from the posterior distribution $\pi$. ABC allows to do parameter inference in cases where the likelihood function is unknown, or too expensive (time-consuming) to evaluate.
It is thus ideal for non-Gaussian observables, where the approximation of a multi-variate Gaussian for the likelihood might be inappropriate. ABC allows us to test this assumption.

The requisite for ABC is a stochastical process that provides simulated realisations \( \{ x_i \} \) of the data \( d \) for a given parameter \( p \) and model class \( M \). These simulations are the model prediction for that parameter; the model is here not a single, deterministic vector \( y(p|M) \) as before, but a stochastic random variable drawn from a distribution.

ABC requires that this distribution is the true underlying probability distribution, or likelihood \( L \), of the observable \( d \). Then, the simulations \( x_i \) are a sample under \( L \), and the density of the sample points represent an estimate of \( L \) of a given parameter \( p \) and model \( M \). For parameter inference, we need to know the probability of the data, \( L(d|p,M) \) given \( p \) and \( M \). that we observed \( d \). How do we get this probability without calculating \( L(d) \)?

This is best illustrated for discrete data: If we have \( N \) model prediction realisations \( \{ x_i \} \) for a parameter \( p \) of some discrete data \( d \), the probability of observing \( d \) is the number of realisations \( r \) for which \( x_i = d \), divided by \( N \). Interestingly, this is a frequentist interpretation of probability, which is the number of matches over the total number of events.

We can write this relative frequency of matches as sum over the distribution (the likelihood \( L \)) of \( x_i \) times the Kronecker delta to only select matches. If in addition the parameter \( p \) is drawn from the prior distribution \( P(p|M) \), the the distribution of matching events, which we call \( \hat{\pi}_{ABC} \), is

\[
\hat{\pi}_{ABC}(p|d,M) := \sum_{i=1}^{N} L(x_i|p,M)P(p|M)\delta_{x_i,d} = L(d|p,M)P(p|M).
\] (61)

This justifies thus the use of the symbol \( \pi \), since the result is an unbiased estimator of the (unnormalised) posterior of \( p \! \)

An additional step is performed with ABC: The number of realisations \( N \) can be as small as one. This might sound surprising at first, since a single realisation \( x \) is not likely to match \( d \), so...
\( \hat{\pi}_{\text{ABC}} \) can be zero. But the probability of a match is equal to \( L(x|p, M) \), independent of the number of realizations \( N \), so the expectation value of (61) is still an unbiased estimator of \( \pi \) even for \( N = 1 \). The large variance of this one-sample estimator is compensated by the large number of sampled parameters \( p \) that are typically explored in sampling of the posterior. It turns out that the overall sampling (or Monte-Carlo) noise does not increase for \( N = 1 \). This one-sample test leads to an accept-reject algorithm.

The application of ABC to continuous data requires further adaptations. Since strict equality between two continuous variables are practically not possible, sampled points \( x_i \) are accepted when they fall within some tolerance level \( \varepsilon \) of \( d \). For multi-variate data, this also requires a metric \( D \) that can be compared to \( \varepsilon \). In addition, the complexity of high-dimensional data is typically reduced to a lower-dimensional space using a summary statistic \( s \) of the data. Thus, a model \( x \) is accepted if

\[
D[s(d), s(x)] < \varepsilon; \quad \text{or equivalently} \quad x \in D_\varepsilon[s(d)],
\]

where \( D_\varepsilon(z) \) is the \( q \)-dimensional ball with radius \( \varepsilon \) centred on \( z \).

The accepted points follow a distribution that is a modified version of (61),

\[
\pi_{\text{ABC,}\varepsilon}(p|d, M) = L_\varepsilon(d|p, M)P(p|M),
\]

where \( L_\varepsilon(d|M) \) is the probability that a proposed parameter \( p \) passes the one-sample tolerance test (62),

\[
L_\varepsilon(d|p, M) = \int d^n x \, L(x|p, M) 1_{D_\varepsilon[s(d)]}(s(x)).
\]

The sum over discrete events \( x_i \) is now an integral over models, and the Kronecker delta has been replaced with the indicator function \( 1_A(x) \), which is unity if \( x \in A \), and zero otherwise. This accept-reject algorithm is illustrated in Fig. 14.

The main assumption of ABC is now that the probability distribution function (64) is a good approximation of the true, underlying likelihood function \( L \) of the data \( d \),

\[
L_\varepsilon(d|p, M) \approx L(d|p, M).
\]

and consequently the ABC posterior (63) an approximation of the true posterior \( \pi \).

This assumption relies on the three ingredients for ABC sampling: the summary statistic \( s \), the distance function \( D \), and the tolerance \( \varepsilon \). Note that traditional Monte-Carlo likelihood sampling approaches also depend on the choice of a summary statistic: In most cases the size of the observed data \( d \) is huge (e.g. CMB pixelised maps or time series, weak-lensing galaxy shape catalogues, SNIa lightcurve time measurements), and is typically reduced to a much smaller observable (e.g. power spectrum, correlation function, magnitude + stretch + color), by a mapping \( s(d) \).

Trevelyan et al. (2009) showed that the choice of a summary statistic and distance function are of great importance for the ABC performance. In Lin & Kilbinger (2015b) we only used one summary statistic, namely \( s(x) = x \), where \( x \) is the (already reduced) data vector. In our case of weak-lensing peak counts (Sect. 2.3.3), \( x \) was chosen to be the number of peaks \( n = (n_1, n_2, \ldots) \) (peak abundance or pdf) as function of peak SNR \( \nu_i \).

Lin et al. (2016) compared two different distance functions \( D \), which are the square root of scalar products of the two vectors \( x \) and \( d \). One distance, \( D_2 \), uses the full data covariance matrix to compute the scalar product, the other one, \( D_1 \), only weighs the data by the variance (diagonal of the covariance). Note that ABC with the distance \( D_1 \) does not necessarily neglect all correlation between data points — these are included automatically in the simulated model predictions under the joint, multi-variate likelihood function \( L \). These correlations are only neglected under \( D_1 \) when deciding whether a model \( x \) for a parameter \( p \) is close enough to \( d \) for \( p \) to be accepted.
Figure 14. Illustration of the ABC accept-reject method in one dimension. Sample points (with positions indicated at the top) are sampled from the underlying likelihood function \( L \) of some observable \( x \). The observed data point marked at \( d \). We want to obtain the likelihood function at the position of the data, \( L(x = d) \). A Monte-Carlo estimate of \( L(d) \) is given by the density of sample points at \( d \). ABC approximates this density by defining a tolerance \( \varepsilon \) around \( d \), and counting the number of points within this limit, normalised by the total number of samples. In a one-sample test limit, this frequency is the acceptance probability, which is equal to the green area divided by the total area under the \( L \) curve.

We found that \( D_1 \) and \( D_2 \) give very similar results when data points are weakly correlated, but \( D_1 \) provided overly tight constraints on parameters for highly correlated data, with under-estimation of the error of \( \Sigma_8 \) by 40%.

The tolerance \( \varepsilon \) is best set in an iterative approach. This automatically solves the problem of fixing the tolerance a priori: If \( \varepsilon \) is too large, too many points are accepted. In that case, \( L_{\varepsilon} \) (64) approaches unity, and ABC effectively provides samples from the prior \( P \). If \( \varepsilon \) is too close to zero, so is \( L_{\varepsilon} \), and sampling becomes very inefficient, with the overwhelming majority of points being rejected. In (Lin & Kilbinger 2015b) we use an iterative importance-sampling algorithm together with ABC. This sequential Monte-Carlo (SMC) ABC method has similarities with population Monte Carlo (Sect. 3.4), and is therefore also called PMC ABC.

4. Measuring weak lensing

The cosmological interpretation of cosmic shear measurements requires the observation of large and deep sky areas in superb image quality, together with sophisticated image analysis methods. A significant detection of cosmic shear requires a very large number of galaxies to high redshifts (of order unity) and low signal-to-noise ratios (down to 10 - 15). The shapes of those faint galaxies have to be measured with high accuracy. Galaxy images have to be corrected for the point-spread function (PSF). The PSF is the combined effect of the imaging system consisting of the atmosphere (for ground-based surveys), telescope optics, and detector. To estimate the PSF, a very pure sample
of stars, uncontaminated by small galaxies, has to be selected. Shape measurements typically have to be calibrated. To ensure measurement biases small enough compared to the statistical errors, large sets of realistic image simulations need to be used. In addition, the interpretation of shape correlations depends crucially on the redshift distribution of the lensed galaxy sample, see (21). Multiple optical band observations have to be used to estimate photometric redshifts. This section gives a brief overview of the methods for weak lensing measurements I have used in my work.

4.1. Galaxy shape measurement

The first family of shape measurement methods is moment-based direct estimation. One of the most widely used methods of this type is KSB (Kaiser, Squires & Broadhurst 1995). KSB measures ellipticity directly on an image using weighted second-order moments of the galaxy brightness distribution. The convolution of the image with the PSF is approximated by linear operations on the ellipticities; it is therefore a perturbative method. Alternatively, instead of correcting the ellipticities as in KSB, the PSF deconvolution can be done directly on the moments of the galaxy light distribution (Rhodes et al. 2000). This is more rigorously explored by DEIMOS (deconvolution of image moments), a truncated hierarchy of higher-order moments in Melchior et al. (2011).

KSB was used in our cosmic-shear measurements (Fu et al. 2008, Schrabback et al. 2010). Deconvolution in moment space, inspired by DEIMOS, but using unweighted moments on denoised images, was one of my contributions to the GREAT3 challenge.

A second large family of shape measurement methods is based on model fitting. These indirect methods assume a model for the surface brightness $I$, including ellipticity parameters, and fit the model to the observed image. One of the advantages of such forward-fitting methods is the straightforward treatment of the PSF: The model is more easily and robustly convolved with the PSF than the observed (pixelised, noisy, maybe partially masked) image is deconvolved. A fully Bayesian forward-fitting method is lensfit, which measures the posterior distribution of ellipticity for galaxies on individual exposures, and combines the results in a Bayesian way without information loss (Miller et al. 2007, Kitching et al. 2008, Miller et al. 2013). A further notable model-fitting method that I have used is gfit (Gentile et al. 2012). lensfit turned out to be the superior method for CFHTLenS. gfit was used as main method by the CEA-EPFL team in the GREAT3 challenge, and is used in on-going work on shear calibration (Pujol et al. 2017), and for the analysis of the Canada-France Imaging Survey (CFIS), see Sect. 6.1.

4.1.1. Shape measurement biases One can make the very general statement that the non-linear dependence of ellipticity estimators on the light distribution in the presence of noise creates a bias, the so-called noise bias. This bias has been investigated for methods based on moments (Hirata et al. 2004) and model fitting (Refregier et al. 2012, Kacprzak et al. 2014).

A further source of bias is related to incorrect assumptions about the light distribution, the so-called model bias (Voigt & Bridle 2010). Not only model-fitting techniques, but also direct methods are still not free from such assumptions: For example, moment-based methods require weight functions whose profile and size should match the observed images, and in fact, direct and indirect methods show very similar biases related to galaxy morphology (Pujol et al. 2017).

Shape biases can be characterised to first approximation by a multiplicative component $m$, and additive term $c$. These bias parameters are given by the relation between observed and true ellipticity (Huterer et al. 2006, Heymans et al. 2006a),

$$
\varepsilon_{i}^{\text{obs}} = (1 + m_i)\varepsilon_{i}^{\text{true}} + c_i; \quad i = 1, 2.
$$

The shear biases $m$ and $c$ are generally functions of galaxy properties and redshift. Current shape
measurement methods provide shear estimates with residual (after calibration) \( m \) at the 1 to few percent level, and \( c \) between 10\(^{-3} \) and 10\(^{-2} \). Typically, the calibration of measured shears is performed using large image simulations (Sect. 4.4). Recent work has been looking into calibration by using the data themselves, the so-called meta-calibration approach (Huff & Mandelbaum 2017, Sheldon & Huff 2017). Future surveys require the accuracy of calibrated shapes to be on the order of 0.1% (Huterer et al. 2006, Massey et al. 2013, Cropper et al. 2013), see Sect. 6.1.

4.2. PSF correction

To estimate the PSF at the position of a galaxy, one has to select stars on the image, measure their shape, and interpolate the resulting PSF to the position of the galaxy. This requires a sample of suitable stars, i.e. without saturated pixels, not hit by cosmic rays, and uncontaminated by galaxies. A common selection criteria is the identification of the stellar locus in a size-magnitude diagram. This is a region of bright and small objects that is relatively well isolated from resolved galaxies and unresolved, dim objects such as very faint galaxies and detection artefacts. Additionally, colour information can be added to classify stars and galaxies.

The PSF (in form of parameters, a pixellised vignette, or a high-resolution model), is then interpolated onto the galaxy position. For ground-based observations in the past, this has usually been done with a two-dimensional polynomial or a rational function. For mosaic multi-CCD cameras, discontinuities between chips are common and have to be accounted for in the PSF model, for example by performing fits on each chip individually (Miller et al. 2013).

4.3. Error modelling and residual systematics

Any weak lensing data analysis must be completed with a robust error modeling. This step is necessary to quantify any residual systematics caused by an imperfect PSF correction, since those residuals can mimic a cosmological signal.

4.3.1. Star-galaxy correlation

The most commonly used approach is a null test of the correlation between the stellar ellipticities \( \varepsilon^\star \) (before PSF correction) and the corrected galaxy shapes \( \varepsilon \). This star-galaxy ellipticity correlation function is defined as

\[
\xi_{\text{sys}} = \langle \varepsilon^\star \varepsilon \rangle. \tag{67}
\]

Note that a non-zero signal on a small region on the sky could come from chance alignments between PSF pattern and a coherent shear from large-scale structure. To interpret measurements of \( \xi_{\text{sys}} \), this cosmic variance contribution needs to be accounted for (Heymans et al. 2012).

4.3.2. PSF model - residual correlation

Two correlation functions quantify the PSF model. These are the auto-correlation of the PSF residuals, \( D_1 \), and the cross-correlation between PSF and PSF residuals, \( D_2 \) (Rowe 2010). These two correlation functions are defined as

\[
D_1(\theta) = \langle (\varepsilon - \varepsilon_m)^\star (\varepsilon - \varepsilon_m) \rangle (\theta);
\]

\[
D_2(\theta) = \langle \varepsilon^\star (\varepsilon - \varepsilon_m) + (\varepsilon - \varepsilon_m)^\star \varepsilon \rangle (\theta). \tag{68}
\]

Here, \( \varepsilon \) is the observed ellipticity of a star, and \( \varepsilon_m \) the PSF model ellipticity\(^\$\). These functions do not only measure the amount of residuals \( (\varepsilon - \varepsilon_m) \), but the spatial correlation of residuals. This can be introduced by a PSF model that does not well represent the spatial variation of the PSF

\(^\$\) In this sub-section, we drop the super-script `\(^\star\)` to denote ellipticity measured for stars; I remind the reader that the different symbol `\(^\star\)` denotes complex conjugation.
over the detector. In case of a perfect PSF model, both correlation functions are expected to vanish identically, \( \langle D_1 \rangle = \langle D_2 \rangle = 0 \).

Rowe (2010) writes the observed ellipticity \( \varepsilon \) of a galaxy as the sum of true ellipticity \( \varepsilon_t \) and noise \( N \),
\[
\varepsilon = \varepsilon_t + N,
\]
and the model ellipticity \( \varepsilon_m \) as sum of true ellipticity \( \varepsilon_t \) and model uncertainty \( m \),
\[
\varepsilon_m = \varepsilon_t + m.
\]
The two correlation functions can then be written as
\[
D_1(\theta) = -\langle m^* N + N^* m \rangle (\theta) + \langle m^* m \rangle (\theta);
\]
\[
D_2(\theta) = -\langle m^* N + N^* m \rangle (\theta) - \langle m^* \varepsilon_t + \varepsilon_t^* m \rangle (\theta).
\]
In particular the second function \( D_2 \) is a useful diagnostic since it is negative definite. In case of over-fitting, the model tends to fit the noise rather the true ellipticity, which creates correlations between the noise \( N \) and the model uncertainty \( m \), and \( D_2 \) becomes significantly negative.

I illustrate such a case in Fig. 15. Non-zero \( D_1 \) and \( D_2 \) indicate PSF residual correlations, in particular on small scales, where the spatial PSF model, a bi-variate polynomial with varying degree, seems to not well fit data. The increasing negative \( D_2 \) with increasing polynomial degree shows cases of over-fitting for the more complex models, indicating that the true PSF variation shows less degrees of freedom.

A second example is discussed below in Sect. 4.4.1.

4.4. Image simulations

Image simulations have been created as collaborative projects within the weak-lensing community, such as the Shear TESting Project (STEP) with the two consecutive blind tests STEP1 (Heymans et al. 2006a) and STEP2 (Massey et al. 2007a). Public challenges like the GRavitational lEnsing Accuracy Testing (GREAT) projects have been launched to reach out to a larger community, in particular computer science, to invite more ideas to tackle the problem of galaxy shape measurement. This contains GREAT08 (Bridle et al. 2009, Bridle et al. 2010), GREAT10 (Kitching et al. 2011, Kitching et al. 2012, Kitching et al. 2013), and GREAT3 (Mandelbaum et al. 2014).

Those collaborative image simulation projects typically started under simple, well-controlled conditions, for example, a constant PSF, constant shear over the field, and analytical galaxy light distributions with high signal-to-noise. They then progressed to more complex and more realistic images, for example galaxy images based on observed HST deep fields. The purpose of those simulations is to test estimates of shear with amplitudes of a few percent to an accuracy at also the percent level. This is typically quantified in terms of multiplicative and additive bias (66). The number of simulated images is necessarily very large, producing hundreds of gigabytes of data.

In the following sections, I briefly discuss some of my past work on galaxy shape measurement, bias quantification, and PSF modelling with image simulations.

4.4.1. GREAT10

Fig. 16 shows the PSF pattern of a GREAT10 star challenge image. Together with Bernhard Riedl, Diploma student under my and Jochen Weller’s supervision (Ludwigs-Maximilians-Universität München), we fit this pattern with bi-variate polynomials in the pixel coordinates \( x \) and \( y \) of varying degree. The diagnostic functions \( D_1 \) and \( D_2 \) are then calculated using the public software ATHENA (Kilbinger et al. 2014).

The diagnostic correlation functions are plotted in Fig. 17 for increasing polynomial degrees 1 (bi-linear), 3 (bi-cubic) and 5. First, below 50 pixels the correlations cannot be measured since
Figure 15. The PSF diagnostic correlation functions $D_1$ (upper four panels) and $D_2$ (lower four panels) as function of angular scale $\theta$, for the Dark Energy Survey (DES) data challenge #5 simulations. The four sub-panels show PSF polynomial models of increasing degree, from 0 (constant) to 3 (cubic). See Sect. 4.3.2.
there are no pairs of stars due to the finite size postage stamps. Second, there is a very significant correlation of residuals up to a few hundred pixels. Clearly, the polynomial is not a good fit to capture those small-scale PSF variations. Third, on scales above 500 pixels, the linear function still displays strong correlations, but the third-order polynomial best fits the data showing the smallest correlations. Increasing the polynomial order to 5 re-introduces correlations at around 500 pixels, hinting to an over-fitting problem.

4.4.2. CFHTLenS The CFHTLenS collaboration created image simulations (1) to calibrate for shear biases (Sect. 4.1.1), and (2) to model residual systematics correlations (Sect. 4.3).

As a cross-check, two sets of simulations were created, using the (modified) code from the GREAT08 and GREAT10 challenges as described in Bridle et al. (2010), and the SkyMaker package (Bertin 2009), respectively. Several important features were added to the simulations compared to the GREAT challenge:

First, the galaxies were modeled as disk+bulge as fitted by lensfit, with ellipticity and size distribution matching the observed data. A mismatch would result in a wrong bias calibration; for example, the STEP and GREAT08/10 simulations did not include the large observed number of small galaxies, and thus the bias of those objects could not be calibrated correctly.

The additive and multiplicative biases were fitted to all simulated galaxies as functions of size and SNR. I propagated the uncertainties of $m$ from the fits to the correlation function covariance. Their contribution turned out to be negligible, see Fig. 18.

Second, each galaxy got assigned a shear from the CFHTLenS “Clone” N-body simulation (Harnois-Déraps et al. 2012). A realistic cosmological shear component is important in the quantitative analysis of systematics (Heymans et al. 2012).

4.4.3. GREAT3 The GREAT3 weak-lensing image challenge (Mandelbaum et al. 2014) was organised by R. Mandelbaum and B. Rowe, and ran from mid-2013 to April 2014. Results were presented and discussed at the final meeting at CMU in Pittsburgh in May 2014 (Mandelbaum
Figure 17. The diagnostic functions $D_1$ (left column) and $D_2$ (right column) as function of pixel units. The three rows from top to bottom show the case of a PSF interpolation model as bi-variate polynomial in $x$ and $y$ of order 1, 3, and 5, respectively. Figure from Bernhard Riedl’s master thesis (unpublished).

et al. 2015). I was part of the “CEA–EPFL” team including from CEA Florent Sureau, Jean-Luc Starck, Fred Maurice Ngolé Mboula, Stéphane Paulin-Henriksson, and from EPFL Marc Gentile and Frédéric Courbin.

I also submitted results under “CEA_denoise”, for which I first denoised the galaxy images by applying the MR_FILTER multi-scale wavelet filter to the images (Starck et al. 2006). I then used SExtractor (Bertin & Arnouts 1996) to measure unweighted second moments of the galaxy light distribution, and corrected for the PSF in moment space, following Rhodes et al. (2000), as mentioned earlier. The results were however not great and could not compete with the leading group of best methods.

The main shape measurement method of our team was the forward-fitting maximum-likelihood
method {\textit{gf}}it. This algorithm is a new version (rewritten in \textsc{python} from scratch) of \textit{gf}it presented in Gentile et al. (2012), which was used in the GREAT10 galaxy challenge. Galaxies are detected by a \textsc{s}ex\textsc{tractor} run, whose output parameters of centroid position, size, and ellipticity are used as first-guess starting point of the model fitting procedure. Galaxy light profiles are modeled as bulge (Sérsic index \( n = 1 + \) exponential (Sérsic \( n = 4 \)) disk, which are concentric and aligned with identical intrinsic ellipticity. The eight parameters centroid, flux, disc flux fraction, bulge radius, disk radius, and ellipticity were fitted. Several minimizers were explored.

\textbf{Weighing and filtering} Often, measured galaxy shapes are weighted depending on various quantities: the galaxy S/R, size, or the best-fit \( \chi^2 \), fit error bar, or confidence in the result. This down-weighs galaxies with uncertain or biased shape estimates and generally improves the results.

Each of the GREAT3 images had 10,000 galaxies, where two pairwise galaxies had the same intrinsic ellipticity rotated by \( 90 \) deg to reduce shape noise. To fully benefit from this noise cancelling scheme, one has an interest to weigh all galaxies equally. However, for some objects no shape can be attributed, for example if the minimizer does not converge. Thus, at the end of the day, we decided to introduce weights for the galaxies with the hope to reduce the biases.

A first, simple weighing scheme was to eliminate galaxies with large fit residuals by setting their weight to zero. I then developed an improved weighting scheme, which I describe in the following.

For a given GREAT3 image, the \textit{gf}it output parameters for each galaxies were used to create new simulations of that image, with properties of noise, PSF, pixellation etc. similar to the input image. \textit{Gf}it was run on this second simulated image, and a PCA decomposition of measured galaxy parameters was performed on the rms ellipticity between input and output, \(|\Delta\epsilon|\). In other words, the galaxies were classified according to ellipticity bias. Using PCA I could identify combinations \( v_i \) of galaxy parameters that contribute most to this bias, and to devise a weighting scheme to downweigh objects for which those parameters correlated with a large bias. To keep the number of parameters reasonably small, I split them up into several sets of parameters, and studied one set at a time. Here,
I will quote results from one such set of parameters (case ‘a’ in (Mandelbaum et al. 2015)), which are the GFFT flux \(\ln(\ln I_0)\), disc radius \(\ln r_{\text{disk}}\), bulge radius \(\ln r_{\text{bulge}}\), disc flux fraction \(f_{\text{disk}}\).

The dependence of the parameters (linear, logarithmic or double-logarithmic) being most significant to quantify large residuals was established by trying out various variants. The weights were determined by plotting \(|\Delta e|\) versus various PCA components \(v_i\) to select a cut-off value \(v_c\) that provided a good separation of objects with low and high bias \(|\Delta e|\).

A density plot in \(v_0\) (the 0th PCA component) and the ellipticity bias \(|\Delta e|\) is show in Fig. 19. There is a clear tail of high-\(|\Delta e|\) objects for small values of \(v_0\), whereas there is a high density of objects with low \(|\Delta e|\) at \(v_0 = 1\ldots2\). The cut-off value \(v_c = 0.6\) is shown by the horizontal line. Thus, objects with PCA component \(v_0 < v_c\) \((v_0 > v_c)\) show generally a strong (weak) bias \(|\Delta e|\). The first category was assigned the weight \(w_{\text{low}} = 0.2\). Objects in the low-bias category kept their weight \(w_{\text{high}} = 1\).

The 0th PCA component is

\[
v_0 = 0.38 \ln(\ln I_0) + 0.59 r_{\text{disk}} + 0.44 \ln r_{\text{bulge}} - 0.28 f_{\text{disk}}
= \ln \left[ (\ln I_0)^{0.38} r_{\text{disk}}^{0.59} r_{\text{bulge}}^{0.44} e^{-0.28 f_{\text{disk}}} \right].
\]

This means that the bias is smaller for objects with larger flux, disk and bulge radius, but smaller disk fraction. The improvement on the overall GREAT3 of this weighting scheme compared to uniform weights was around 15 - 20%.

4. Redshift estimation

4.5.1. Photometric redshifts Weak lensing observables, being integrals along the line of sight weighted by the source galaxy distribution \(n(z)\) (23), require knowledge of the latter if they are to be interpreted cosmologically. To first order, the mean redshift \(\bar{z}\) has to be determined, but also the shape of \(n(z)\) plays an important role. Huterer et al. (2006) find a rough estimate of \(P_\kappa(\ell \sim 1000) \propto \Omega_{\text{de}}^{-3.5} \sigma_8^{2.9} \bar{z}^{1.6} |w|^{0.31}\). Clearly, for a desired accuracy on cosmological parameters, the mean redshift of sources has to be known to at least that accuracy, and to a much higher accuracy in the case of parameter on which the power spectrum has a weaker dependence such as \(w\).

Spectroscopy of all the faint galaxies used for a typical weak-lensing survey is too costly, and redshifts have to be estimated from broad-band photometry, using the technique of photometric redshifts, or photo-zs. There are various methods to measure photometric redshifts, of which
template-based approaches are one of the most popular. They perform $\chi^2$-type fits of (redshifted) template SEDs to the flux in the observed bands. Exemplary methods that have been used in a weak-lensing context include LePhare (Ilbert et al. 2006) and Bayesian Photometric redshift estimation (BPZ; Benítez 2000). LePhare photo-z’s for CFHTLS were published in (Coupon et al. 2009); BPZ was the method of choice for CFHTLenS (Hildebrandt et al. 2012). Both methods do not only performe a point-estimate $\hat{z}$ but provide the full pdf of the redshift.

4.5.2. Clustering-based redshift estimation

To assess the quality of photometric redshifts and to recover the true redshift distribution, one can make use of the spatial clustering of galaxies. From the amount of cross-correlation of photometric samples between different redshift bins one can deduce the amount of redshift outliers. In Benjamin et al. (2010) we introduced this method, and applied it to photometric clustering (Coupon et al. 2012, McCracken et al. 2015) and weak-lensing (Benjamin et al. 2013).

Fig. 21 is an example from Coupon et al. (2012). For the full galaxy sample of the CFHTLS-Wide T0006 release, it shows the angular auto-correlation functions of redshift bin #1, $0.4 < z < 0.6$, and the two auto-correlation functions of that bin with galaxies at higher redshift bins #3 ($0.8 < z < 1$) and #4 ($1 < z < 1.2$), respectively. From these measurement we estimate the contamination $f_{ij}$ between redshift bin pairs $i$ and $j$ due to photometric mis-identification. We carry out the so-called global pair-wide analysis, considering only two bins at a time while in turn setting to zero contaminations of other bins, and also neglecting higher-order effects such as magnification bias. The resulting constraints of the contamination fractions, $f_{ij}$, $i,j = 1, 3, 4$, are shown in the in the middle and right panels. Around three percent of galaxies in bin #3 are mis-identified into bin #1, which leads to the non-zero cross-correlation between the two bins (dotted line in the left panel). The
mixing between bins #1 and #4 is consistent with zero, as is the cross-correlation function (dashed line in the left panel).

Similarly, the cross-correlation of the photometric with a spectroscopic sample can reveal the true redshift distribution (Newman 2008), although this reconstruction is hampered by a possible redshift-dependent bias of the photometric galaxy sample (Schulz 2010). In Scottez et al. (2016) we applied the cross-correlation method to CFHTLS/VIPERS, and showed that it can yield redshifts for individual galaxies if color information is present.

5. Observational results and cosmological constraints

This section highlights some of the observational results for cosmological parameter constraints from cosmic shear to which I contributed.

5.1. Second-order statistics

In the following sub-sections I discuss the basic observational results from second-order cosmic shear statistics. The parameter combination that cosmic shear (including non-linear scales) is most sensitive to the parameter $\Sigma_8$, defined by

$$\Sigma_8 = \sigma_8 \left( \frac{\Omega_m}{\Omega_{m,0}} \right)^\alpha,$$

with typical values of $\alpha \approx 0.5 - 0.7$. The pivot value $\Omega_{m,0}$ can be chosen freely. Fig. 22 and Table 3 shows this combination measured in recent years for $\Omega_{m,0} = 0.3$. If the original measurement (indicated in the table) corresponds to a different pivot value, I transform to $\Omega_{m,0} = 0.3$ including a simple error propagation computation.

5.1.1. CFHTLS-T0003, 2007 The third data release (T0003) of the wide part of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) with an observed area of 53 deg$^2$ provided 2D cosmic shear results out to very large, linear scales (7.7 degrees, corresponding to 170 Mpc at the mean lens redshift of 0.5; Fu et al. 2008). This enabled us to infer cosmological constraints using large scales only, thereby reducing uncertainties from non-linear and baryonic physics on small scales. Using $\langle M^2_{ap}(\theta) \rangle$ on scales $\theta > 85'$ we obtained $\sigma_8(\Omega_m/0.25)^{0.53} = 0.837 \pm 0.084$. This study used the
photometric redshifts from the 4 square degree deep part of CFHTLS (Ilbert et al. 2006), taking into account sampling variance. The deep fields have an area of 4 square degrees, an increase of nearly a factor 2500 over the HDF. Constraints on neutrino masses were obtained using this data release in Tereno et al. (2009).

By that time, ground-based surveys had become large enough to enable detailed residual systematics tests. For CFHTLS, Kilbinger et al. (2009) as well as first multi-colour observations (Fu 2008) revealed an anomalous shear amplitude scaling with source redshift and a variance between MegaCam pointings larger than expected. Kilbinger et al. (2009) quantified their influence on cosmological parameters from a joint analysis. However, it took three more years of work by the CFHTLenS team (Sect. 5.1.3) of around 20 members and a complete re-analysis of CFHTLS images to finally obtain a robust shear catalogue. The origin of those systematics was never found, and the price to pay for a systematic-free data set was to reject 25% of the MegaCam pointings that were plagued with PSF residuals.

5.1.2. Re-analysis of COSMOS, 2009 I contributed to a re-analysis of the weak-lensing data (Schrabback et al. 2010), independent from the first series of papers from that survey (Leauthaud et al. 2007, Massey et al. 2007b). We obtained improved photo-z’s from twice the number of bands (Ilbert et al. 2009). Due to the low number of high-S/N stars in ACS fields, and temporal instabilities of HST, the PSF model was obtained by PCA of the PSF pattern from dense stellar fields. In this work we presented a five-bin tomographic analysis, leading to constraints on the deceleration parameter $q_0 = -\ddot{a}/a^2 = \Omega_m/2 - \Omega_\Lambda$, with a detection of acceleration ($q_0 < 0$) at 94.3% confidence, including additional priors on $h$ and $\Omega_b$.

Table 3: Parameter values of $\sigma_8\Omega_m^{\alpha}$ used for Fig. 22. For different surveys (first column) the second column is the value rescaled to $\Omega_m = 0.3$ which is plotted in the figure, obtained from the original measurement (third column). The fourth column indicates the number of redshift bins, the column 5 is the reference. This table is an updated version of Table 1 from Kilbinger (2015).

| Survey       | $\sigma_8(\Omega_m/0.3)^a$ | original measurement | $N_z$ | Reference                        |
|--------------|-----------------------------|----------------------|-------|----------------------------------|
| COSMOS       | 0.810 ± 0.17                | $\sigma_8(\Omega_m/0.3)^{0.48} = 0.81 \pm 0.17$ | 1     | Massey et al. (2007b)           |
| COSMOS       | 0.866$^{+0.085}_{-0.068}$    | $\sigma_8(\Omega_m/0.3)^{0.44} = 0.866^{+0.085}_{-0.068}$ | 3     | Massey et al. (2007b)           |
| 100 deg$^2$  | 0.736 ± 0.0438              | $\sigma_8(\Omega_m/0.24)^{0.59} = 0.84 \pm 0.05$ | 1     | Benjamin et al. (2007)          |
| CFHTLS       | 0.699 ± 0.0383              | $\sigma_8(\Omega_m/0.25)^{0.64} = 0.785 \pm 0.043$ | 1     | Fu et al. (2008)                |
| CFHTLS       | 0.760 ± 0.0763              | $\sigma_8(\Omega_m/0.25)^{0.53} = 0.837 \pm 0.084$ | 1     | Fu et al. (2008), large scales  |
| COSMOS       | 0.750 ± 0.08                | $\sigma_8(\Omega_m/0.3)^{0.51} = 0.75 \pm 0.08$ | 5+1$^a$ | Schrabback et al. (2010)        |
| COSMOS       | 0.650 ± 0.1                 | $\sigma_8(\Omega_m/0.3)^{0.62} = 0.65 \pm 0.1$ | 1     | Schrabback et al. (2010)        |
| COSMOS       | 0.69$^{+0.08}_{-0.14}$      | $\sigma_8(\Omega_m/0.3)^{0.46} = 0.69^{+0.08}_{-0.14}$ | 1     | Semboloni et al. (2011), +IA    |
| SDSS-Stripe82| 0.585$^{+0.074}_{-0.112}$   | $\sigma_8(\Omega_m/1.0)^{0.7} = 0.252^{+0.012}_{-0.052}$ | 1     | Lin et al. (2012)              |
| SDSS-Stripe82| 0.597$^{+0.11}_{-0.15}$     | $\sigma_8(\Omega_m/0.264)^{0.67} = 0.65^{+0.12}_{-0.15}$ | 1     | Huff et al. (2014)             |
| SDSS-DR7     | 0.721 ± 0.0451              | $\sigma_8(\Omega_m/0.25)^{0.57} = 0.8 \pm 0.05$ | 1     | Mandelbaum et al. (2013)        |

$^a$ 5 narrow photo-z bins and one wide bin of faint galaxies.
$^b$ 3D lensing, no z-binning.
$^c$ The index $\alpha = 0.5$ is adopted from the WMAP9 measurement, published on http://lambda.gsfc.nasa.gov.
| Survey         | obs(Ωm/0.265)^{0.5} | obs(Ωm/0.27)^{0.59} | obs(Ωm/0.27)^{0.65} | obs(Ωm/0.27)^{0.55} | obs(Ωm/0.27)^{0.46} | obs(Ωm/0.27)^{0.44} | ref                     |
|----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|-------------------------|
| DLS            | 0.816 ± 0.0385       | 0.868 ± 0.041        | 1                    |                |
| CFHTLenS       | 0.740 ± 0.0301       | 0.787 ± 0.032        | 1                    |                |
| CFHTLenS       | 0.738±0.0654         | 0.79±0.067           | 1                    |                |
| CFHTLenS       | 0.728 ± 0.0377       | 0.771 ± 0.04         | 2                    |                |
| CFHTLenS       | 0.737±0.0305         | 0.774±0.032          | 6                    |                |
| CFHTLenS       | 0.657 ± 0.21         | 0.69 ± 0.22          | \(\infty\)^b        |                |
| CFHTLenS       | 1.107 ± 0.258        | 1.16 ± 0.27          | \(\infty\)^b        |                |
| CFHTLenS       | 0.725±0.0694         | 0.77±0.071           | 1                    | Fu et al. (2014), +IA |
| CFHTLenS       | 0.785±0.028          | 0.84±0.031           | 1                    | Liu et al. (2015a)  |
| SDSS-Stripe82  | 0.785 ± 0.0287       | 0.82 ± 0.03          | 1                    | Liu et al. (2015b)  |
| DES-SV         | 0.811±0.059          | 0.811±0.059          | 3                    | The Dark Energy Survey Collaboration et al. (2016) |
| DLS            | 0.818±0.0344         | 0.818±0.0343         | 5                    |                |
| CFHTLenS       | 0.738 ± 0.0393       | 0.404 ± 0.0215       | 7                    |                |
| DES-SV         | 0.770 ± 0.07         | 0.77 ± 0.07          | 1                    |                |
| CFHTLenS       | 0.670 ± 0.03         | 0.67 ± 0.03          | 2                    |                |
| KiDS           | 0.745 ± 0.039        | 0.745 ± 0.039        | 4                    |                |
| KiDS           | 0.801 ± 0.0324       | 0.801 ± 0.0324       | 4                    |                |
| KiDS           | 0.742 ± 0.035        | 0.742 ± 0.035        | 4                    |                |
| KiDS           | 0.696±0.048          | 0.696±0.048          | 1                    |                |
| DES-Y1         | 0.789±0.024          | 0.789±0.024          | 4                    |                |
| DES-Y1         | 0.783±0.024          | 0.783±0.021          | 4                    |                |
| KiDS           | 0.750 ± 0.059        | 0.75 ± 0.059         | 1                    |                |
| DES-Y1         | 0.78±0.051           | 0.78±0.051           | 4-5                  |                |
| WMAP3^c        | 0.671 ± 0.0669       | 0.76 ± 0.05          | -                    | Spergel et al. (2007) |
| WMAP5^c        | 0.738 ± 0.0544       | 0.796 ± 0.036        | -                    | Komatsu et al. (2009) |
| WMAP7^c        | 0.774±0.0512         | 0.814±0.031          | -                    | Komatsu et al. (2011) |
| WMAP9          | 0.792 ± 0.0529       | 0.434 ± 0.029        | -                    | Hinshaw et al. (2013) |
| Planck2013     | 0.848 ± 0.0286       | 0.89 ± 0.03          | -                    | Planck Coll. (2014a), C(ℓ) |
| Planck2013     | 0.740 ± 0.0242       | 0.764 ± 0.025        | -                    | Planck Coll. (2014b), SZ |
| Planck2015     | 0.825 ± 0.0161       | 0.4521 ± 0.0088      | -                    | Planck Collaboration et al. (2016), C(ℓ) |
| Planck2015+SPT | 0.808 ± 0.0324       | 0.598 ± 0.024        | -                    | Simard et al. (2018) |

### 5.1.3. CFHTLenS, 2012 - 2015

A milestone for cosmic shear represented the CFHT lensing survey (CFHTLenS; Erben et al. 2013). With CFHTLenS we provided measurements that relied on independently cross-checked photometric redshifts, and a robust estimate of residual systematics
Figure 22. Mean and 68% error bars for the parameter $\sigma_8 (\Omega_m/0.3)^{0.5}$, for various cosmic shear observations, plotted as function of their publication date (first arXiv submission). All parameter values are given in Table 3. Different surveys are distinguished by colour as indicated in the figure. Data points are shown for second-order statistics (circles), third-order (diamonds), 3D lensing (pentagons), galaxy-galaxy lensing (+ galaxy clustering; triangle), and CMB (squares). This plot is an updated version of Fig. 7 from Kilbinger (2015).

on weak-lensing shear correlations (Sect. 4.3). In these studies, the residual systematics analysis was done completely independently from the cosmological parameters analysis, in order not to bias the cosmological results.

Photometric redshifts for each source galaxy were obtained in Hildebrandt et al. (2012), the robustness of which was verified using spectroscopic redshifts, COSMOS 30-band photo-zs, and a cross-correlation analysis (Benjamin et al. 2013). Galaxy shapes were measured on individual exposures with lensfit and calibrated using two independent suites of image simulations (Miller et al. 2013). An excess correlation between star and galaxy shapes (67) was found on 25% of the observed fields, which in turn were discarded from the cosmological analysis (Heymans et al. 2012). Two-dimensional cosmic shear correlation functions from CFHTLenS were presented in Kilbinger et al. (2013). A two-bin tomographic analysis was performed by Benjamin et al. (2013). The same tomographic data were used to place constraints on modified gravity (Simpson et al. 2013). Further, a six-bin tomographic analysis was performed where cosmological and intrinsic-alignment parameters were constrained simultaneously (Heymans et al. 2013). Late-type galaxies were found to not show
any significant intrinsic alignment, while for early type galaxies IA was detected at about 2σ.

For a $\Lambda$CDM cosmology, cosmic shear constrains a combination of $\Omega_m$ and $\sigma_8$ that is perpendicular to the one obtained from CMB (Contaldi et al. 2003). Adding cosmic shear to WMAP (Wilkinson Microwave Anisotropy Probe) results in typical reduction of error bars on $\Omega_m$ and $\sigma_8$ of up to 50%, similar to other low-$z$ cosmological probes such as Baryonic Acoustic Oscillations (BAO). For example, the WMAP7 constraints of $\Omega_m = 0.273 \pm 0.03$ and $\sigma_8 = 0.811 \pm 0.031$ (Komatsu et al. 2011) get tightened when adding CFHTLenS, resulting in $\Omega_m = 0.274 \pm 0.013$ and $\sigma_8 = 0.815 \pm 0.016$.
Planck’s cosmological findings from temperature anisotropies (together with CMB lensing and WMAP polarization) correspond to a higher matter density and normalization compared to most previous probes, with $\Omega_m = 0.315 \pm 0.017$ and $\sigma_8 = 0.829 \pm 0.012$, or $\sigma_8(\Omega_m/0.27)^{0.46} = 0.89 \pm 0.03$ (Planck Coll. 2014a). This is consistent with CFHTLenS at the $2\sigma$ level, see Fig. 25. Further, Planck’s counts of Sunyaev-Zel’dovich (SZ) clusters results in a lower normalization of $\sigma_8(\Omega_m/0.27)^{0.3} = 0.78 \pm 0.01$ (Planck Coll. 2014b). Sect. 5.1.4 discusses whether adding extra-parameters such as massive neutrinos are needed to reconcile recent high- and low-z data.

A model with variable curvature does not change the cosmic-shear constraints on $\Omega_m$ and $\sigma_8$ by a lot. Pre-Planck CMB data alone cannot constrain the curvature of the Universe, and adding other probes such as measurements of $H_0$ or weak lensing are required. Planck and high-resolution ground-based millimetre-wavelength radio telescopes of similar sensitivity and resolution such as SPT (South Pole Telescope) and ACT (Atacama Cosmology Telescope) have detected weak-lensing of the CMB by large-scale structures (CMB lensing), which helps to break the geometrical degeneracy. This results in tight constraints on $\Omega_K$ from CMB alone (Sherwin et al. 2011, van Engelen et al. 2012, Planck Coll. 2014a). Fig. 25 shows joint cosmic shear and CMB constraints for a free-curvature model.

Since the effect of dark energy on the suppression of the growth of structure is relatively small, 2D weak lensing is not very sensitive to dark energy. Tomographic weak lensing can place interesting constraints on the dark-energy equation of state parameter. Fig. 24 shows how CMB constraints from WMAP7 — with an additional prior on $H_0$ from Riess et al. (2011) — are reduced by CFHTLenS six-bin tomography. The parameters $\Omega_m$ and $w_0$ are measured to better than 10% accuracy, for both a flat and free-curvature $wCDM$ model. The improvement is similar to adding Baryonic Acoustic Oscillation (BAO) data from the SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS; Anderson et al. 2012) to CMB data.

Constraints on modified gravity using the parametrization in eqs. (6, 7) showed consistency with GR (Simpson et al. 2013). A simple model was considered where $\Sigma$ and $\mu$ did not vary spatially, and at early times tend towards GR, so that deviations of GR are allowed at late times where the accelerated expansion happens. The present-day values of those two parameters were measured to be $\Sigma_0 = 0.00 \pm 0.14$, and $\mu_0 = 0.05 \pm 0.25$, combining CFHTLenS weak-lensing tomographic data (Benjamin et al. 2013), redshift-space distortions from WiggleZ (Blake et al. 2012) and 6dFGS (Beutler et al. 2012), WMAP7 CMB anisotropies from small scales, $\ell \geq 100$ (Larson et al. 2011), and the Riess et al. (2011) $H_0$ prior (see Fig. 26).

5.1.4. Further results from CFHTLenS Other groups different from the CFHTLenS collaboration have used those data for further, extended analysis. The cosmological constraints from CMB temperature anisotropies measured by the Planck satellite (Planck Coll. 2014a) seem to be in slight tension with other probes. In particular, Planck found a higher power-spectrum normalisation $\sigma_8$. Several works proposed massive neutrinos to alleviate the tension with low-z probes such as weak lensing: Massive neutrinos are still relativistic at recombination and do not significantly influence the CMB anisotropies. They become however non-relativistic at late time, and dampen the growth of structure, therefore reducing the low-z clustering power. Joint analyses including massive neutrinos from Planck and our CFHTLenS weak-lensing data of Kilbinger et al. (2013) were found to improve parameter constraints with the inclusion of non-zero neutrino masses (Battye & Moss 2014, Beutler et al. 2014), but the evidence still favors a $\Lambda$CDM model without additional parameters for massive neutrinos (Leistedt et al. 2014).
Figure 25. Cosmic shear and CMB 68.3% and 95.5% confidence levels for \( \Omega_m \) and \( \sigma_8 \) in a \( \Lambda \text{CDM} \) universe. (a) Assuming flatness. CFHTLenS 2D, 6-bin tomography, 3D, and 3D from large scales only are compared to Planck constraints. (b) With free curvature, showing CFHTLenS (joint second- and third-order), WMAP9, Planck, CFHTLenS + WMAP9, and CFHTLenS + Planck constraints, from Fu et al. (2014).

Figure 26. Combined constraints on the present-day modified-gravity parameters \( \Sigma_0 \) and \( \mu_0 \), from redshift-space distortions (RSD), cosmic shear (CFHTLenS), and their combination, including the case of additional BAO (Anderson et al. 2012) and large-scale WMAP7 (ISW) data. All data are combined with a \( H_0 \) prior and small-scale CMB data, see text. From Simpson et al. (2013).
5.2. Higher-order correlations

The motivation behind higher-order shear statistics has been argued for in Sect. 2.3. Even though the measurement is challenging and the overall signal-to-noise ratio is low, several significant detections of third-order shear correlations have been made.

Several higher-order measurements resulted from CFHTLenS. First, van Waerbeke et al. (2013) measured the skewness of reconstructed convergence maps and found good agreement with WMAP7 predictions. After validating the data for shear residual third-order correlations, Simon et al. (2015) performed a cosmological analysis of the aperture-mass skewness exploring a non-Gaussian likelihood. Fu et al. (2014) combined the second- and third-order aperture-mass combined with WMAP9 and Planck to obtain cosmological results (see Fig. 25), including models of intrinsic alignment and source-lens clustering as astrophysical systematics.

Quite a few measurements of weak-lensing peak counts were obtained using data from the CFHT/MegaCam Stripe-82 survey (CS82), CFHTLenS, the Dark Energy Survey (DES), and the Kilo-Degree Survey (KiDS) (Shan et al. 2014, Liu et al. 2015a, Liu et al. 2015b, Kacprzak et al. 2016, Martinet et al. 2018). Preliminary measurements on the combination of DES (SVD; science verification data), KiDS data release DR1/2, and CFHTLenS, and comparison to our fast stochastic models are presented and discussed in Chieh-An Lin’s PhD thesis (Lin 2016).

5.3. Intrinsic alignment

In Fu et al. (2008) we tried to obtain constraints on intrinsic alignment. Since we had no redshift information and therefore one large redshift bin, the different contributions to the shape correlation function of cosmic shear ($G\theta$) and shear-shape intrinsic alignment ($GI$) are not easily separated. The shape-shape alignment ($II$) cannot be resolved with a wide redshift distribution. We therefore obtained a non-detection of $GI$, with $A = 2.2^{+3.6}_{-4.6} \times 10^{-7} h/$Mpc, using the simple model of Heymans et al. (2006b). No evidence pointed to the presence of $GI$ above the statistical and systematic error level.

A 2σ detection of intrinsic alignment from early-type galaxies was obtained by jointly fitting cosmology and the Hirata et al. (2004) linear IA model to CFHTLenS cosmic shear tomographic data (Heymans et al. 2013). Fig. 27 shows a systematic lower amplitude of shear correlation for cross-redshift correlations compared to the cosmic-shear prediction, as expected from a negative $GI$ contribution. No detection was found for the late-type sample.

6. Future cosmic shear expectations and forecasts

6.1. Upcoming and future surveys

New instruments, either cameras, telescopes, or both, are being designed and built specifically for the purpose of weak-lensing observations. Their design is driven by the goal to provide superb image quality with very small, uniform, and well-understood image distortions. The pixel scale is chosen to sufficiently sample the PSF. In view of the enormous costs of new experiments, in particular space missions, the instruments have to be thoroughly and carefully designed to guarantee the desired scientific outcome, for example, the measurement of dark-energy properties with a given accuracy.

In early 2017, CFIS, the Canada-France Imaging Survey will start observations. CFIS is attributed around 200 nights over three years, to map the Northern sky in $u$ and $r$. The $u$-band part (CFIS-LUAU) will complement the on-going LUAU program, and cover at the end 10,000 square degree, with a limiting magnitude of $u = 24.4$. The $r$-band survey, CFIS-WIQD (for Wide + Image
Figure 27. Amplitudes of tomographic measures of $\xi_+$ at $\theta = 1$ arcmin for redshift bins ($ij$), against peak lensing efficiency redshift $z_{\text{peak}}$ for early-type galaxies, from CFHTLenS. The free parameters $\alpha^\text{fid}$ multiplied with a WMAP7 fiducial $GG$ model $\xi_{+,\text{fid}}$ were fitted to $\xi_+$ and $\xi_-$, simultaneous for all redshift bins and angular scales. At low $z$, the auto-correlations ($i = j$, open circles) lie above the fiducial dashed line ($\alpha^\text{fid} = 1$), as expected for a $II > 0$ contribution. The cross-correlations ($i \neq j$, filled circles) lie systematically below the prediction, indicating a $GI < 0$ contamination. From Heymans et al. (2013).

Quality + Deep) will observe 4,800 deg$^2$ north of $\delta > 30^\circ$ and galactic latitude $b > 25^\circ$, to a depth of $r = 24.85$.

Despite only observing in two bands (with a rather shallow $u$ component), CFIS will be very interesting for weak lensing, in particular in combination with deep spectroscopic surveys such as eBOSS and DESI. Such deep data does not exist in the Southern hemisphere. CFIS will also contribute to DESI target selection and provide photometric bands for Euclid photo-z’s.

Going to space offers the two major advantages: Escaping atmospheric turbulence leads to a stable and small PSF, and infrared observations provide photo-z’s to significantly higher redshifts than from the ground.

The accessible extra-galactic sky (the area outside the Milky Way and the ecliptic) of 15,000 deg$^2$ will be observed from space with the ESA satellite mission Euclid (Laureijs et al. 2011). The two main science drivers for Euclid are cosmic shear and galaxy clustering, which will be observed using three instruments, an optical imager, a near-infrared imager, and a near-infrared slitless spectrograph. The optical imager on board Euclid is designed to have a very stable PSF, both spatially as well as in the time domain. To collect enough galaxy light from billions of high-redshift galaxies (30 arcmin$^{-2}$), the transmission curve is very broad, corresponding to the combined $R + I + z$ filters, with a required depth of $R + I + z = 24.5$. This poses new challenges to overcome, in particular galaxy colour gradients and PSF calibrations.

Further obstacles unique to space-based observations will have to be tackled (Cropper et al. 2013): For example, the very small PSF will be undersampled by the pixels of size 0.1 arcsec. From these undersampled stellar images, a reliable, high-resolution PSF model has to be reconstructed. Furthermore, the detector degrades with time due to the bombardment with cosmic rays, and the shapes of objects get distorted by the so-called charge transfer inefficiency (CTI). Corrections as function of time, position on chip, and brightness of the objects have to be applied...
The shear calibration of the huge data expected from Euclid to the required formidable accuracy will be another challenge. The necessary volume of image simulations for calibration is estimated to be huge, as well as the processing time of those simulations (Hoekstra et al. 2017). It is not clear yet whether new, alternative methods such as meta-calibration, can be used: The undersampled galaxy images with very large wavelength range makes deblending with the correct PSF challenging, and will require the development of new methods. We are studying sparsity-approaches (Farrens et al. 2017) that we plan to develop further for Euclid.

6.2. Outlook

In 2000, cosmic shear was first measured over a few square degrees of observed sky, from some ten thousand galaxies. Fifteen years later, surveys have increased these numbers by a factor of 100, imaging a few million galaxies over $O(100)$ square degrees. Many challenges were met to analyse these data, taking years of work. This resulted in constraints on cosmological parameters that are competitive compared to other cosmological probes.

In another ten years, upcoming and future experiments will cover a substantial fraction of the entire sky, measuring billions of galaxies. This signifies yet another data volume increase of a factor of 100, not to mention the data quality improvement due to instruments dedicated to weak lensing. The formidable challenge here is reducing systematic errors to an acceptable level when analysing these large data sets. New, unprecedented difficulties have to be overcome, for example CTI for Euclid, and blended galaxy images for LSST. To fully exploit those surveys, large follow-up programs are needed to obtain the necessary large samples of photometric and spectroscopic redshifts. In addition, to interpret the results of those surveys, the accuracy of theoretical predictions of the non-linear power spectrum including baryonic physics need to be significantly improved.

If all these challenges can be overcome, weak cosmological lensing has the great potential to advance our understanding of fundamental physics. It can explore the origin of the recent accelerated expansion of the Universe, and distinguish between dark energy models and theories of modified gravity. Cosmic shear can measure initial conditions of the primordial Universe, constrain the mass of neutrinos, and measure properties of dark matter. Not only that, the study of intrinsic galaxy alignments has provided insights into the formation and evolution of high-redshift galaxies in their dark-matter environment, proving that cosmic shear does not only probe cosmology, but influences and enriches other areas of astrophysics. Thus, over the last fifteen years, weak cosmological lensing has established itself as a major tool in understanding our Universe, and with upcoming large surveys, it will continue to be of great value for astrophysics and cosmology.

Acknowledgments

I would like to thank the members of the jury for agreeing to be in the committee of my habilitation à diriger des recherches (HDR). I am very grateful to Alain Blanchard, Martin Kunz, and Christophe Pichon, for the thoroughness of their reports on this thesis, and thus on my work over the last ten or so years. I would also like to thank Jim Bartlett and Nick Kaiser for their roles as examiners, and Stéphane Plaszczynski for being the jury chairman. For their useful comments on earlier versions of this manuscript I would like to thank Alain Blanchard and Stéphane Plaszczynski.

I would also like to thank Bernhard Riedl, whose master thesis work was the basis for some of the unpublished material in Section 4.4.
Although a pre-requisite for the HDR, I find supervising students a real pleasure. Much of the work presented here was made possible by those students. These include Tim Eifler, Liping Fu, Jean Coupon, Christopher Bonnett, Laura Wolz, Bernhard Riedl, and Chieh-An Lin, for whom I would like to thank.

I am indebted to my colleagues at the astrophysics department (DAp) of CEA Saclay for their continuous help and support over the last seven years. Thank you, Pierre-Olivier Lagage, Michel Talvard, Anne Decourchelle, Pascale Delbourgo, Jérôme Rodriguez, Marc Sauvage, Jérôme Amiaux, Marguerite Pierre, Koryo Okumura, Pierre-Antoine Frugier, and Eméric LeFloc’h, as well as everyone I had the pleasure to work with at the CosmoStat laboratory, Jean-Luc Starck, Jérôme Bobin, Florent Sureau, Sandrine Pires, Stéphane Paulin-Henriksson, Valeria Pettorino, Austin Peel, François Lanusse, Fred Ngolé Mboula, Arnau Pujol, Santiago Casas, Kostas Themelis, Axel Guinot, and Morgan Schmitz.

Finally, I thank my parents for their support, Melissa for her love and care, and Léon for introducing a new element of chaos, creativity, and neverending inspiration to my life.

References

Albrecht A, Bernstein G, Cahn R, Freedman W L, Hewitt J & al. 2006 arXiv 0609591.
Alsing J, Heavens A & Jaffe A H 2017 MNRAS 466, 3272–3292.
Amendola L, Kunz M & Sapone D 2008 JCAP 4, 13.
Anderson L, Aubourg E, Bailey S, Bizyaev D, Blanton M & al. 2012 MNRAS 427, 3435–3467.
Anderson T W 2003 An introduction to multivariate statistical analysis third edn Wiley-Interscience.
Asgari M, Schneider P & Simon P 2012 A&A 542, A122.
Bacon D J, Réfrégier A R & Ellis R S 2000 MNRAS 318, 625–640.
Bartelmann M 2010 Classical and Quantum Gravity 27(23), 233001–+.
Bartelmann M & Schneider P 2001 Phys. Rep. 340(4-5), 297–472.
Battye R A & Moss A 2014 Physical Review Letters 112(5), 051303.
Benabed K & Scoccimarro R 2006 A&A 456, 421–432.
Benítez N 2000 ApJ 536, 571–583.
Benjamin J, Heymans C, Semboloni E, Van Waerbeke L, Hoekstra H & al. 2007 MNRAS 381, 702–712.
Benjamin J, van Waerbeke L, Heymans C, Kilbinger M, Erben T & al. 2013 MNRAS 431, 1547–1564.
Benjamin J, van Waerbeke L, Ménard B & Kilbinger M 2010 MNRAS 408, 1168–1180.
Berger J & Jeffreys W 1992 Amer. Statistician 80, 64–72.
Bernardeau F 1998 A&AS 133, 375–382.
Bernardeau F, Bonvin C, Van de Rijt N & Vernizzi F 2012 Phys. Rev. D 86(2), 023001.
Bernardeau F, Bonvin C & Vernizzi F 2010 Phys. Rev. D 81(8), 083002.
Bernardeau F, Van Waerbeke L & Mellier Y 1997 A&A 322, 1–18.
Bernstein G M 2009 ApJ 695, 652–665.
Bertin E 2009 Mem. Soc. Astron. Italiana 80, 422.
Bertin E & Arnouts S 1996 A&AS 117, 393–404.
Beutler F, Blake C, Colless M, Jones D H, Staveley-Smith L & al. 2012 MNRAS 423, 3430–3444.
Beutler F, Saito S, Brownstein J R, Chuang C H, Cuesta A J & al. 2014 MNRAS 444, 3501–3516.
Blake C, Brough S, Colless M, Contreras C, Couch W & al. 2012 MNRAS 425, 405–414.
Blandford R & Narayan R 1986 ApJ 310, 568–582.
Bridle S, Balan S T, Bethge M, Gentile M, Harmeling S & al. 2010 MNRAS 405, 2044–2061.
Bridle S, Shawe-Taylor J, Amara A, Applegate D, Balan S T & al. 2009 Annals of Applied Statistics 3, 6–37.
Calabrese E, de Putter R, Huterer D, Linder E V & Melchiorri A 2011 Phys. Rev. D 83(2), 023011.
Cameron E & Pettitt A N 2012 MNRAS 425, 44–65.
Cappé O, Douc R, Guillin A, Marin J M & Robert C 2008 Statist. Comput. 18(4), 447–459.
Cappé O, Guillin A, Marin J M & Robert C 2004 J. Comput. Graph. Statist. 13(4), 907–929.
Capranico F, Merkel P M & Schäfer B M 2013 MNRAS 435, 194–206.
Chevallier M & Polarski D 2001 International Journal of Modern Physics D 10, 213–223.
Clifton T, Ferreira P G, Padilla A & Skordis C 2012 Phys. Rep. 513, 1–189.
Coles P & Lucchin F 1996 Cosmology: The Origin and Evolution of Cosmic Structure John Wiley & Sons.
Contaldi C R, Hoekstra H & Lewis A 2003 Physical Review Letters 90(22), 221303/1–4.
Cooray A & Hu W 2001 ApJ 548, 7–18.
Cooray A & Sheth R 2002 Phys. Rep. 372, 1–129.
Coupon J, Ilbert O, Kilbinger M, McCracken H J, Mellier Y & al. 2009 A&A 500, 981–998.
Coupon J, Kilbinger M, McCracken H J, Ilbert O, Arnouts S & al. 2012 A&A 542, A5.
Covone G, Sereno M, Kilbinger M & Cardone V F 2014 ApJ 784, L25.
Crittenden R G, Natarajan P, Pen U L & Theuns T 2002 ApJ 568, 20–27.
Cropper M, Hoekstra H, Kitching T, Massey R, Amiaux J & al. 2013 MNRAS 431, 3103–3126.
de Felice A & Tsujikawa S 2010 Living Reviews in Relativity 13, 3.
DES Coll., Abbott T M C, Abdalla F B, Alarcon A, Aleksic J & al. 2017 arXiv 1708.01530.
Dietrich J P & Hartlap J 2010 MNRAS 402, 1049–1058.
Dodelson S 2003 Modern cosmology Amsterdam (Netherlands): Academic Press.
Dodelson S & Schneider M D 2013 Phys. Rev. D 88(6), 063537.
Dodelson S, Shapiro C & White M 2006 Phys. Rev. D 73(2), 023009–+.
Dodelson S & Zhang P 2005 Phys. Rev. D 72(8), 083001.
Dvai G, Gabadadze G & Porrati M 2000 Physics Letters B 484, 112–118.
Dyson F W, Eddington A S & Davidson C 1920 Royal Soc. of London Philosoph. Transactions Series A 220, 291–333.
Efstathiou G 2008 MNRAS 388, 1314–1320.
Eifler T, Kilbinger M & Schneider P 2008 A&A 482, 9–19.
Eifler T, Schneider P & Hartlap J 2009 A&A 502, 721–731.
Eifler T, Schneider P & Krause E 2010 A&A 510, A7.
Einstein A 1916 Annalen der Physik 354, 769–822.
Eisenstein D J, Zehavi I, Hogg D W, Scoccimarro R, Blanton M R & al. 2005 ApJ 633, 560–574.
Erben T, Hildebrandt H, Miller L, van Waerbeke L, Heymans C & al. 2013 MNRAS 433, 2545–2563.
Evrard A E 1989 ApJ 341, L71–L74.
Fan Z, Shan H & Liu J 2010 ApJ 719, 1408–1420.
Farrens S, Ngol` e Mboula F M & Starck J L 2017 MNRAS 482, 502–514.
Erben T, Kilbinger M & Schneider P 2010 MNRAS 401, 1264–1274.
Fu L, Kilbinger M, Erben T, Heymans C, Hildebrandt H & al. 2014 MNRAS 441, 2725–2743.
Fu L, Semboloni E, Hoekstra H, Kilbinger M, van Waerbeke L & al. 2008 A&A 479, 9–25.
Futamase T 2015 International Journal of Modern Physics D 24, 1530011.
Gentile M, Courbin F & Meylan G 2012 arXiv:1211.1447.
Giannantonio T, Porciani C, Carron J, Amara A & Pillepich A 2012 MNRAS 422, 2854–2877.
Gruen D, Friedrich O, Krause E, DeRose J, Cawthon R & al. 2017 arXiv 1710.05045.
Hamana T 2001 MNRAS 326, 326–332.
Hamana T, Colombi S T, Thion A, Devriendt J E G T, Mellier Y & al. 2002 MNRAS 330, 365–377.
Hamilton A J S 2000 MNRAS 312, 257–284.
Hamilton A J S, Rimes C D & Scoccimarro R 2006 MNRAS 371, 1188–1204.
Harnois-D´ eraps J, Vafaei S & van Waerbeke L 2012 MNRAS 426, 1262–1279.
Hartlap J, Schrabback T, Simon P & Schneider P 2009 A&A 504, 689–703.
Hartlap J, Simon P & Schneider P 2007 A&A 464, 399–404.
Hastings W K 1970 Biometrika 57, 97–109.
Heymans C, Grocutt E, Heavens A, Kilbinger M, Kitching T D & al. 2013 MNRAS 432, 2433–2453.
Heymans C & Heavens A 2003 A&A 399, 307.
Heymans C, Van Waerbeke L, Bacon D, Berge J, Bernstein G & al. 2006a MNRAS 368, 1323–1339.
Heymans C, Van Waerbeke L, Miller L, Erben T, Hildebrandt H & al. 2012 MNRAS 427, 146–166.
Heymans C, White M, Heavens A, Vale C & van Waerbeke L 2006b MNRAS 371, 750–760.
Hilbert S, Marian L, Smith R E & Desjacques V 2012 MNRAS 426, 2870–2888.
Hildebrandt H, Erben T, Kuijken K, van Waerbeke L, Heymans C & al. 2012 MNRAS 421, 2355–2367.
Hildebrandt H, Viola M, Heymans C, Joudaki S, Kuijken K & al. 2017 MNRAS 465, 1454–1498.
Hinshaw G, Larson D, Komatsu E, Spergel D N, Bennett C L & al. 2013 ApJS 208, 19.
Hinshaw G, Weiland J L, Hill R S, Odegard N, Larson D & al. 2009 ApJS 180, 225–245.
 Hirata C M, Mandelbaum R, Seljak U, Guzik J, Padmanabhan N & al. 2004 MNRAS 353, 529–549.
Hoekstra H & Jain B 2008 Annual Review of Nuclear and Particle Science 58, 99–123.
Hoekstra H, Viola M & Herbonnet R 2017 MNRAS 468, 3295–3311.
