A new approach to photometric redshift contamination, providing critical insight for weak lensing cosmology.

by

Jonathan Remby Embro Benjamin

B.Sc Physics, Bishop's University, 2004

M.Sc Physics, University of British Columbia, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Astronomy)

The University of British Columbia
(Vancouver)

August 2013

Abstract

Light travelling through the Universe is deflected by the presence of mass, this effect will distort the size and shape of observed galaxies. Weak gravitational lensing measures the amount of mass in the Universe by observing these subtle changes in the shapes of distant galaxies. In order to properly interpret the observed shapes of galaxies their distances must be accurately known, this information is encoded in the redshift distribution.

A detailed spectroscopic observation is the most reliable way to measure the redshift of a galaxy. Unfortunately this is a time-intensive process and weak lensing surveys are composed of millions of galaxies many of which are too faint for spectroscopic observation. For this reason photometric redshifts are used. Photometric redshifts are less accurate than spectroscopic redshifts but are easier to obtain since they rely on only a few measurements over large ranges of wavelength.

Thorough knowledge of uncertainties in the photometric redshifts is vital to weak lensing because photometric redshifts provide the distances necessary to understand the weak lensing signal. In this thesis we present a new technique to measure the reliability of photometric redshifts with the goal of improving the estimated redshift distribution for use in weak lensing studies. Mock observational surveys are used to test the technique before applying it to two surveys: the Deep component of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS-Deep) and the CFHT Lensing Survey (CFHTLenS). We demonstrate our ability to construct both the true redshift distribution and the true average redshift of galaxies in a given photometric redshift range. Furthermore, we show that the photometric redshift probability distribution function can be used as an accurate measure of the true redshift distribution when summed for an ensemble of galaxies.

Using our tested redshift distribution we present cosmological constraints for CFHTLenS from a weak lensing analysis. We present constraints on cosmological parameters for a model of the Universe with dark energy and cold (non-interacting) dark matter (ΛCDM). We find that our weak lensing analysis, combined with other cosmological probes, improves the precision of these measurements by a factor of 1.5 to 2.

Preface

A version of Chapter 2 has been published: J. Benjamin, L. Van Waerbeke, B. Ménard, and M. Kilbinger, *Photometric redshifts: estimating their contamination and distribution using clustering information*, 2010, MNRAS[1], Volume 408, Issue 2, pp.1168-1180. The author of this thesis was the primary investigator and was solely responsible for authoring the paper. The collaborators provided valuable discussions in the formulation and development of the research and in critiquing the written work. This work also resulted in the author contributing a related analysis in another published work: T. Erben, H. Hildebrandt, M. Lerchster, P. Hudelot, J. Benjamin, L. Van Waerbeke, T. Schrabback, F. Brimioulle, O. Cordes, J. P. Dietrich, K. Holhjem, M. Schirmer, P. Schneider, *CARS: the CFHTLS-Archive-Research Survey. I. Five-band multi-colour data from 37 sq. deg. CFHTLS-wide observations*, Astronomy and Astrophysics, Volume 493, Issue 3, 2009, pp.1197-1222. In this case the thesis author was responsible for the analysis and writing contained in Section 4.1.1. with the exception of Figure 10 which was produced by Hendrik Hildebrandt.

A version of Chapter 3 has been published: J. Benjamin, L. Van Waerbeke, C. Heymans, M. Kilbinger, T. Erben, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, B. Rowe, T. Schrabback, F. Simpson, J. Coupon, L. Fu, J. Harnois-Déraps, M. J. Hudson, K. Kuijken, E. Semboloni, S. Vafaei, M. Velander, *CFHTLenS tomographic weak lensing: Quantifying accurate redshift distributions*, 2013, MNRAS, Volume 431, Issue 2, pp.1547-1564. The author of this thesis was the primary investigator and was solely responsible for authoring the paper with the exception of Section 3.4 which was first drafted by Catherine Heymans. All collaborators provided comments on the manuscript, and participated in discussions concerning the direction of the research.

[1]Monthly Notices of the Royal Astronomical Society

Table of contents

List of tables

List of figures

Glossary

ACF Angular correlation function. The two-point correlation function of galaxies on the sky. The ACF is positive if the amount of clustering is in excess of what is expected from a random distribution of points.

BAO Baryon Acoustic Oscillations. In the early Universe baryons and photos were coupled. When they decoupled a characteristic scale was imprinted on the baryonic density fluctuations. BAO studies attempt to recover this information by looking at the distribution of galaxies.

BOSS Baryon Oscillation Spectroscopic Survey. A survey designed to study BAOs and described in Anderson et al. (2012).

BPZ Bayesian Photometric redshift. A publicly available code used to measure photometric redshifts. BPZ uses Bayesian statistical methods to estimate a posterior PDF for each galaxy's photometric redshift.

CARS CFHTLS Archive Research Survey. A data-mining survey carried out by Erben et al. (2009).

CCD Charge Coupled Device. A silicone chip technology capable of electronic imaging. CCDs and related technologies make digital cameras possible.

CDM Cold dark matter. There is evidence that dark matter has an extremely small cross-section for self-interaction. Such fluids are referred to as 'cold'. Most cosmological models assume that dark matter is cold.

CFHT Canada-France-Hawaii Telescope. CFHT is an 8 meter class telescope with a 1 square degree camera known as MEGACAM. It uses 5 optical filters u^*, g', r', i', and z'.

CFHTLenS CFHT Lensing Survey. Applies data-reduction and data-processing techniques tailored to the study of weak lensing to the CFHTLS-Wide data.

CFHTLS	CFHT Legacy Survey. A large survey conducted by CFHT. Composed of a deep component CFHTLS-Deep and a wide component CFHTLS-Wide.
CFHTLS-Deep	The deep component of the CFHTLS survey. CFHTLS-Deep is composed of 4 fields each subtending 1 square degree on the sky.
CFHTLS-Wide	The wide component of the CFHTLS survey. CFHTLS-Wide is composed of 4 fields with varying sky coverage. The total sky area observed is 179 square degrees.
CMB	Cosmic Microwave Background. Primordial photons which have been free-streaming since they decoupled from the matter in the early Universe. As such they are a direct probe of the early Universe.
COSMOS	Cosmic Evolution Survey. COSMOS was observed with the Hubble space telescope's Advanced Camera for Surveys. COSMOS is a deep survey covering 1.4 square degrees on the sky.
Euclid	Euclid is a planned space-based mission which plans to survey 15,000 square degrees with optical and infrared photometric bands from 550nm to 200nm.
FLRW	The Friedmann-Lemaître-Robertson-Walker metric can be derived from assumptions of homogeneity and isotropy. It relates the interval between events in space-time (ds) to intervals in one or both of time (dt) and 3-dimensional space ($d\mathbf{\Sigma}$)
HST	Hubble Space Telescope. The HST distance ladder is a suit of observations meant to establish cosmic distances in order to measure the Hubble constant with high precision (Riess et al., 2011).
ΛCDM	The standard model of cosmology which requires a cosmological constant Λ (identified as dark energy) and cold dark matter.
LSS	Large Scale Structure. The arrangement of matter on the large scales in the Universe. N-body simulations such as the Millennium Simulation predict that dark matter is arranged in filamentary structures. The Universe is homogeneous only on very large scales $\sim$ 10 Mpc.
LSST	Large Synoptic Survey Telescope. A planned ground based 8 meter class telescope. It will image 30,000 square degrees in 6 photometric bands from 320nm to 1050nm.

MS	Millennium Simulation. A large N-body simulation starting from small density perturbations and evolving them forward in time simulating the formation of large scale structure (LSS) in the Universe.
PDF	Probability Distribution Function. In this thesis the photometric redshift PDFs describe the probability that a galaxy has a given redshift.
Photometric redshift	A redshift estimate from broadband photometric observations. Observed broadband magnitudes are compared to template spectral energy distributions convolved with the transmission curves of the braodband filters. Less accurate than spectroscopic redshifts but they require less observing time making them essential for large surveys.
Planck	An all-sky survey to measure the cosmic microwave background over the whole sky. It significantly improves the resolution of the earlier survey conducted by WMAP.
R11	Constraints on the Hubble constant from the HST distance ladder as presented in Riess et al. (2011).
SED	Spectral Energy Distribution. The energy distribution as a function of wavelength (or frequency) of light emitted from a galaxy. If individual emission or absorption features can be identified the redshift of the galaxy can be determined.
SNAP	SuperNova Acceleration Probe. A planned space telescope optimized to detect Supernovae and conduct a deep wide field survey ideal for weak gravitational lensing. The survey component will map $\sim 10,000$ square degrees in 6 photometric bands from 350nm to 1700nm.
Spectroscopic redshift	Very precise redshift measurements that require long observing times. Individual emission or absorption features are identified and their shift in wavelength is attributed to cosmological redshift.
VVDS	VIRMOS VLT Deep Survey. VIRMOS is the VIsible imaging Multi-Object Spectrograph allowing up to 600 simultaneous spectra to be obtained. VLT is the Very Large Telescope located at the European Southern Observatory. VVDS is a deep imaging and redshift survey of more than 150,000 redshifts covering 16 square degrees.
Weak gravitational lensing	The distortion of light from background galaxies by intervening mass. The distortions have an isotropic term κ, the convergence, which changes the observed size of the galaxy, and an anisotropic term γ, the shear, which induces an observed ellipticity.

WMAP Wilkinson Microwave Anisotropy Probe. WMAP is a space mission to measure the CMB temperature anisotropies over the entire sky. Data obtained after seven years of observation (and the measured cosmological parameters) are referred to as WMAP7.

Symbols

$a(t)$ Scale factor. Defined to be unity today $t = 0$ and $a(t) < 1$ in the past.

c Speed of light in a vacuum.

χ Comoving distance. Due to the expansion of the Universe proper distance is a function of time. The Comoving distance is defined such that it is constant with time.

f_{ij} The contamination fraction. The number of galaxies which originate in redshift bin i but are found to be in redshift bin j divided by the true (uncontaminated) number of galaxies in redshift bin i.

$f_K(\chi)$ Comoving angular diameter distance.

γ The shear. The anisotropic distortion caused by weak gravitational lensing which causes a galaxy to have an induced ellipticity.

H_0 Hubble constant. The slope of the linear relationship between the distance to a galaxy and the rate at which it is moving away from us $v = H_0 r$. Typically given in the units of $km\,s^{-1}\,Mpc^{-1}$.

h The dimensionless Hubble constant $h = \frac{H_0}{100\mathrm{km\,s^{-1}\,Mpc^{-1}}}$.

κ The convergence. Describes the isotropic distortion caused by weak gravitational lensing which changes the observed size of a galaxy.

n_{s} Slope of the primordial mass density power spectrum.

Ω_{b} Energy density of baryons.

Ω_{m} Energy density of matter (baryons + dark matter).

$\boldsymbol{\Omega_\Lambda}$	Energy density of dark energy.
$\boldsymbol{P_\delta}$	The mass density power spectrum. Quantifies on which scales mass over-densities exist.
$\boldsymbol{P_\kappa}$	The convergence power spectrum. Quantifies on which scales weak gravitational lensing correlates the shapes of galaxies.
$\boldsymbol{\sigma_8}$	Normalisation of the mass density power spectrum.
$\boldsymbol{w_0}$	Constant term in the dark energy equation of state, $w(a) = w_0$.
$\boldsymbol{\omega_{\rm ij}(\theta)}$	The angular correlation function. This describes the excess probability, compared to a random distribution, of finding pairs of galaxies between samples i and j at separation θ. If i=j this is the auto-correlation function, if not it is the cross-correlation function. In this thesis i and j represent different redshift bins.

Chapter 1

Introduction[1]

Cosmology is the study of the large scale properties of the Universe, including its origin, evolution, and ultimate fate. Cosmology as a physical science is a relatively young field, which began sometime after Albert Einstein published the theory of General Relativity in 1917. At that time, the existence of other galaxies was controversial. The large scale implications of the theory of General Relativity were beyond the scope of the scientific establishment.

The scale of our perceived Universe increased dramatically in the subsequent dozen years. Edwin Hubble made two historic observations that challenged our understanding of the cosmos. First, in 1923, Hubble made observations of novae, bright stellar events with well characterised brightness, in the Milky Way's closest neighbour, the Andromeda galaxy. The distances measured from these observations established that Andromeda was not within the Milky Way, but was an analogous system with trillions of stars. Hubble's second historic observation came in 1929, when he demonstrated that galaxies more distant from the Milky Way were receding at increased speeds.

Georges Lemaître had already published a theoretical study in 1927 that used the equations of General Relativity to describe an expanding fabric of space-time, which could explain Hubble's observations. The equations of General Relativity depend on the amount of mass and other forms of energy in the Universe. Large amounts of mass would mean that the expansion of space would eventually stop and reverse. With small amounts of mass, the expansion of space would continue indefinitely. The content of the Universe is key to understanding which of these scenarios will come to pass.

After conducting detailed observations of Milky Way stars, Jan Oort concluded in 1932 that the visible matter in our Galaxy could not account for the orbital motions observed. There was a significant undetected mass component. Dubbed dark matter, this form of matter does not emit or absorb light, making it impossible to detect except by the gravitational force it exerts on visible matter. The influence of dark matter in galaxy clusters was observed by Fritz Zwicky in 1933, suggesting that dark matter was the dominant gravitational mass in all locations and over all scales in the Universe. Modern observations estimate that about 85 per cent of the mass in the Universe is in the form of dark matter (for e.g., Dodelson, 2003). Therefore, the amount of dark matter in the Universe is critical in cosmology.

[1]Versions of Sections 1.2 and 1.3 (excluding 1.3.4) of this chapter have been published. J. Benjamin. Cosmological constraints from the 100 square degree weak lensing survey, 2007, M.Sc. Thesis, University of British Columbia, Canada

The final pillar of modern cosmology became well established in the late 1990s through the observations of distant supernovae by two groups: Riess et al. (1998) and Perlmutter et al. (1999). It was discovered that the expansion of the Universe is accelerating. The easiest way to explain this in General Relativity is through the cosmological constant, Λ, which acts as a repulsive form of energy. This is referred to as dark energy due to its unintuitive behaviour.

Modern observations indicate that dark energy accounts for about 70 per cent of the energy density of the Universe. Matter accounts for about 30 per cent, and of that 30 per cent only about 4 per cent is ordinary matter, and the rest is dark matter (for e.g., Dodelson, 2003). This model of the Universe is called Λ Cold Dark Matter (ΛCDM) and has been extremely successful at explaining observations. However, many mysteries remain. We have yet to determine the true nature of either dark matter or dark energy, despite our ability to measure their impact on cosmological observations. Cosmologists' contribution on this frontier is to increase the precision with which we can test the ΛCDM model as well as alternative models. Precise measurements can be used to rule out certain particle physics candidates for dark matter, theoretical models for dark energy, and modifications to General Relativity. Cosmological observations may hold the key to unlocking the mysteries of the so-called dark sector of the Universe.

Future survey missions such as the Large Synoptic Survey Telescope (LSST), the SuperNova Acceleration Probe (SNAP), and Euclid[2] are predicted to measure the contributions from dark matter and dark energy to unprecedented precision. A central observational tool in these missions is weak gravitational lensing, which is distinguished by its ability to detect dark matter and ordinary matter simultaneously through the gravitational distortion of light from distant galaxies. With new levels of precision in observations comes the need for a more precise theoretical understanding. In order to take full advantage of weak lensing from these future surveys, work must be done to improve our ability to model the weak lensing signal in order to properly interpret these observations. This thesis presents a novel technique to improve our understanding of the distances to galaxies. This is critical in the modelling of weak lensing and the future of precision cosmology.

The following sections cover some fundamental background theory and concepts from cosmology, weak gravitational lensing, and photometric redshifts. These three topics are relevant to the research presented in this thesis. In the broadest context, the work presented in this thesis is important because it improves our understanding of the content of the Universe. Specifically, this work improves weak gravitational lensing measurements which are critical to understanding both dark matter and dark energy.

1.1 Cosmology

This section introduces some essential equations and parameters that will be used in this thesis. For more comprehensive works on cosmology, the reader may be interested in the following textbooks: Ryden (2003), Dodelson (2003) and Weinberg (2008). The goal here is to introduce key equations and parameters in order to gain some insight into their importance and physical meaning.

[2]See the glossary for more details of these missions.

1.1.1 Proper and comoving distances

Distance measures are extremely important to cosmology. The fact that the Universe is expanding means that distances must be carefully defined.

The Friedmann-Lemaître-Robertson-Walker (FLRW) metric can be derived from assumptions of homogeneity and isotropy. It relates the interval between points (more generally called events) in space-time (ds) to intervals in one or both of time (dt) and 3-dimensional space ($d\Sigma$):

$$ds^2 = -c^2dt^2 + a(t)^2 d\Sigma^2, \tag{1.1}$$

where $a(t)$ is the scale factor. The scale factor is defined such that $a(t_{\rm now}) = 1$ is today. At earlier times, the Universe had expanded less and the scale factor was less than unity. In this way, $a(t)$ encodes the expansion history of the Universe. Most cosmological models have $0 < a(t) \leq 1$ for $t \leq t_{\rm now}$. It is convenient to use spherical coordinates and to write the 3-dimensional term with an explicit dependence on the curvature of space:

$$d\Sigma^2 = d\chi^2 + f_K(\chi)^2(d\theta^2 + sin^2\theta d\phi^2), \tag{1.2}$$

where χ is the radial comoving distance (which will be defined in a moment), θ and ϕ are the polar and azimuthal angles respectively, and $f_K(\chi)$ is a piecewise defined factor that depends on the curvature of space (K):

$$f_K(\chi) = \begin{cases} \frac{\sin(\chi\sqrt{K})}{\sqrt{K}} & \text{for } K > 0 \\ \chi & \text{for } K = 0 \\ \frac{\sinh(\chi\sqrt{|K|})}{\sqrt{|K|}} & \text{for } K < 0. \end{cases} \tag{1.3}$$

The curvature is given by $K > 0$ for an open universe, $K < 0$ for a closed universe, and $K = 0$ for a familiar flat Euclidean space. The 3-dimensional distance between two points at a given time $dt = 0$ (i.e. the distance one would measure if one could lay out measuring sticks between two locations) is called the proper distance, and is given by $d_p = a(t)\int d\Sigma$. If we consider the distance from Earth to some distant object, the proper distance has only a radial component: $d_p = a(t)\chi$. The comoving distance is defined as $d_p/a(t)$; this distance is constant for all values of the scale factor (i.e. it does not change with time). In the purely radial case, the comoving distance is χ.

Consider a yardstick of known proper length L, which subtends an angle of $\Delta\theta$. The distance to such a yardstick can be deduced from the angle it subtends. For small angles ($\Delta\theta << 1$), the angular diameter distance is defined by:

$$d_A \equiv \frac{L}{\Delta\theta}. \tag{1.4}$$

If we label the two ends of the yardstick θ_1 and θ_2, then $\Delta\theta = |\theta_2 - \theta_1|$. If we use the FLRW metric, the interval between the two ends of the yardstick is

$$ds = a(t)\, f_K(\chi)\Delta\theta. \tag{1.5}$$

Identifying ds as the length of the yardstick L gives us an expression for the proper angular diameter

distance,

$$d_A = a(t) f_K(\chi). \tag{1.6}$$

In analogy to the proper and comoving distances, the term $f_K(\chi)$ can be understood as the comoving angular diameter distance. For flat space, the comoving angular diameter distance is equal to the comoving distance χ, and the proper angular diameter distance is equal to the proper distance $a(t)\chi$.

1.1.2 Cosmological parameters

In 1929, Edwin Hubble observed that galaxies are moving away from Earth (and the Milky Way) with radial velocities (v), which are linearly dependent on their proper distance, r, from Earth,

$$v = H_0 r, \tag{1.7}$$

where H_0 is Hubble's constant today. Using a typical value from contemporary literature and conventional units,

$$H_0 = 70\,\mathrm{km\,s^{-1}Mpc^{-1}}, \tag{1.8}$$

where Mpc represents a megaparsec, which is equal to $\sim 3.09 \times 10^{19}$km. For convenience, the dimensionless Hubble constant is defined to be

$$h \equiv \frac{H_0}{100\mathrm{km\,s^{-1}Mpc^{-1}}}. \tag{1.9}$$

More generally, it is possible to relate a time dependent Hubble parameter to the scale factor and its time derivative ($\dot{a}$),

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \tag{1.10}$$

The evolution of the Hubble parameter is governed by the Friedmann equation. Conceptually, this equation establishes how the Universe will expand (or contract) as a function of time given the energy density of its constituents. The Friedmann equation can be written as:

$$H(t)^2 = H_0^2 \left[\frac{\Omega_\mathrm{m}}{a(t)^3} + \Omega_\Lambda + \frac{1-\Omega_0}{a(t)^2} \right], \tag{1.11}$$

where the current energy density of matter is given by Ω_m, the current energy density of dark energy is given by Ω_Λ, and the current total energy density is $\Omega_0 = \Omega_\mathrm{m} + \Omega_\Lambda$. For a flat Universe, $\Omega_0 = 1$, the open and closed cases are described by $\Omega_0 > 1$ and $\Omega_0 < 1$ respectively. The Ω parameters describe the energy density of each component as a fraction of the current critical energy density. The critical energy density is the average energy density of the Universe required for spatial flatness. Taking the energy density of matter as an example,

$$\Omega_\mathrm{m} = \frac{E_{\mathrm{m},0}}{E_{\mathrm{crit},0}} = E_{\mathrm{m},0} \frac{8\pi G}{3c^2 H_0^2}, \tag{1.12}$$

where G is Newton's gravitational constant and c is the speed of light.

Cosmological probes such as weak gravitational lensing allow us to measure these cosmological parameters. The values of these parameters are integral to understanding the past, present, and future evolution of the Universe. The subsequent sections will show how weak gravitational lensing studies are capable of putting constraints on cosmology.

1.2 Gravitational lensing

Gravitational lensing – the deflection of a light bundle's trajectory by the presence of mass – is one of the most fundamental results of general relativity. The phenomenon was experimentally verified in 1919 during a solar eclipse in which background stars were seen to have shifted with respect to their usual positions due to the gravity of the Sun. This experiment, carried out by Sir Arthur Eddington (Eddington, 1920), served as a resounding affirmation of Einstein's theory of general relativity.

The deflection of star light due to the Sun is very small, owing to the small mass of the Sun and the relative distances of the observer-Sun-star system. The strength of a gravitational lens – its ability to alter the light coming from background sources – depends on the mass of the lens and the ratio of distances,

$$\alpha \propto \frac{D_{ol} D_{ls}}{D_{os}}, \tag{1.13}$$

where α is the angular deflection of the light, D_{ol} is the distance between the observer and the lens, D_{os} is the distance between the observer and the source, and D_{ls} is the distance between the lens and the source. Thus the effect is quite small for the Sun lensing background stars, since $D_{os} \simeq D_{ls}$ and D_{ol} is relatively small.

Little work was done on the theory of gravitational lensing at that time. Chwolson (1924) published a work wherein he considered a perfectly co-aligned source and foreground mass, concluding that the resulting image would be a ring around the lensing mass. Einstein (1936) wrote a paper considering the lensing effects by stars. He derived equations for image locations, separation, and magnification, and concluded that the separation of the images would be far too small (on the order of milliarcseconds) to be resolved. Inspired by Einstein, Zwicky published two papers in 1937 (Zwicky 1937a, 1937b), which focused on the potential lensing signal from extragalactic nebulae (i.e. galaxies). He concluded that the separation of images would be about 10 arcseconds and hence easily resolved. In addition, the magnification of the source would allow the observation of distant galaxies. He calculated the probability of a distant source galaxy being lensed, concluding that about 1 in 400 would be affected, practically guaranteeing the existence of gravitationally lensed galaxies.

The field lay dormant until the 1960s, when several works (Klimov, 1963; Liebes, 1964; Refsdal, 1964a,b) extended the theory of gravitational lensing and outlined its usefulness to astronomy. When the first quasars were detected in 1963, it was realised that they were far more distant than most galaxies and therefore could be lensed by the galaxies. However, it was not until 1979 that the first multiply imaged quasar was detected by Walsh et al. (1979) with a redshift of $z{\sim}1.4$.

Aided by the recent development of the Charge Coupled Device (CCD), gravitational lensing saw much advancement during the 1980s. Of particular interest here is the first detection of giant luminous

arcs in galaxy clusters (Lynds & Petrosian, 1986; Soucail et al., 1987). These faint arcs are stretched tangentially with respect to the cluster center, extending about 10 times further in this direction than in the radial direction. These arcs were later identified as highly distorted and magnified images of background galaxies (Mellier et al., 1991; Paczynski, 1987), whose light was being strongly lensed by the gravitational field of the galaxy cluster. Less distorted arcs than these were named arclets (Fort et al., 1988), and are visible in many galaxy clusters.

Gravitational lensing has become a large field with many branches of research, including strong lensing, micro lensing, galaxy-galaxy lensing, and weak lensing. This thesis is concerned only with the last application; for more detailed reviews of gravitational lensing, the reader is referred to Schneider et al. (1992) and Petters et al. (2001).

1.3 Weak gravitational lensing

1.3.1 Weak lensing by galaxy clusters

Unlike giant luminous arcs, or the less pronounced arclets, galaxies affected by weak gravitational lensing do not exhibit an immediately identifiable distortion. That is, the distortion of their ellipticity is not easily distinguishable from the intrinsic ellipticity. However, if the distortions vary slowly with position then nearby galaxies will be distorted in a similar way. Assuming that there is no correlation of the intrinsic ellipticities, a local ensemble average of galaxy ellipticities will provide a measure of the distortion due to weak lensing. This signal was first detected in 1990 around two galaxy clusters (Tyson et al., 1990). The affect of intrinsic alignments of galaxy ellipticities is discussed below in Section 1.3.4.

One of the most appealing aspects of weak lensing is its ability to measure mass in an unbiased way; it is not sensitive to the form of matter (baryonic or dark matter) or its state. Strong lensing can also be used to measure the mass of clusters, provided that there are enough giant luminous arcs present. However, weak lensing can probe the matter distribution to much larger radii and can be applied to clusters that do not have a strong lensing signal. The theoretical edifice of weak lensing by clusters was detailed by Kaiser & Squires (1993), wherein it was shown that the measurement of galaxy distortions can be used to construct a parameter-free map of the two-dimensional projected mass distribution. The first mass reconstruction of a cluster was carried out a year later by Fahlman et al. (1994).

For a review of weak lensing by galaxy clusters, the reader is referred to Fort & Mellier (1994).

1.3.2 Weak lensing by large scale structure

Arguably the most remarkable application of weak lensing is the measurement of mass on cosmic scales. Light travelling through the inhomogeneous large scale structure (LSS) of the Universe is deflected, causing shape distortions of background galaxies, aptly named ‘cosmic shear’. A statistical treatment of cosmic shear reveals details of the matter distribution of the Universe, which can be compared with reliable theoretical models of structure growth. Hence, it can be used to constrain cosmology.

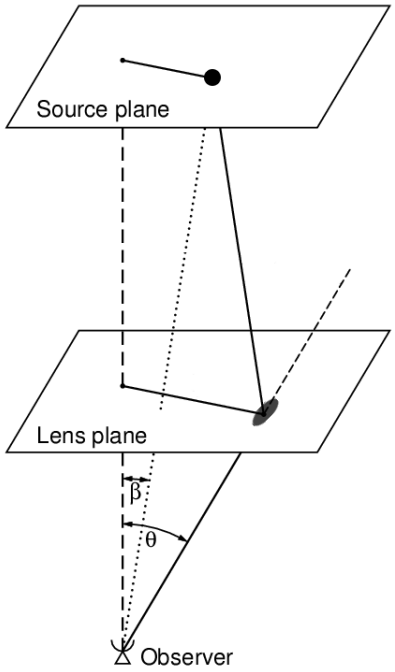

Figure 1.1: Sketch of the lensing geometry, the dashed line is the optical axis defined by the line connecting the observer and the center of the lensing mass. The circular source in the source plane has a position described by β, its lensed image in the lens plane has a position described by θ. (Source: Bartelmann & Schneider (2001))

The first mention of light deflection by LSS is often credited to Gunn (1967). However, like many great ideas it can be traced back to Richard Feynman, specifically a lecture given by him at Caltech in 1964[3]. The theory of light propagation in an inhomogeneous universe and the development of weak lensing as a statistical treatment of galaxy distortions were explored by several theorists (e.g., Babul & Lee, 1991; Blandford et al., 1991; Jaroszynski et al., 1990; Kaiser, 1992; Kristian & Sachs, 1966; Lee & Paczynski, 1990; Schneider & Weiss, 1988; Villumsen, 1996). Theoretical studies concerned with the measurement of cosmological parameters via weak lensing were soon to follow (Bernardeau et al., 1997; Hu & Tegmark, 1999; Jain & Seljak, 1997; Kaiser, 1998; Kamionkowski et al., 1998; Van Waerbeke et al., 1999). Finally, after decades of theoretical study, weak lensing by LSS was observationally detected by several groups (Bacon et al., 2000; Kaiser et al., 2000; Van Waerbeke et al., 2000; Wittman et al., 2000). The following section will briefly introduce the concepts of weak lensing. For more detailed reviews, the reader is referred to Bartelmann & Schneider (2001), Van Waerbeke & Mellier (2003), and Schneider (2005).

1.3.3 Cosmological dependence

A light bundle observed at a position θ has been deflected by the LSS of the Universe and has a position in the source plane given by β (see Figure 1.1). The distortion of images can then be described by the

[3]Refregier (2003), who cites a personal communication with J.E. Gunn

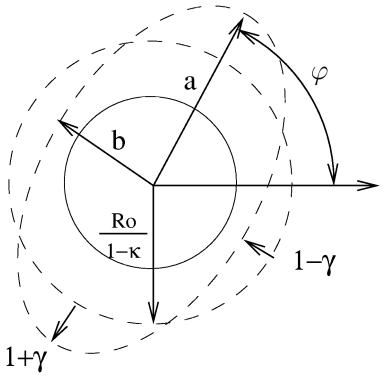

Figure 1.2: Schematic of the first order effects weak lensing has on a circular background galaxy of radius R_0. The convergence is an isotropic distortion that increases the radius, scaling R_0 by a factor of $(1-\kappa)^{-1}$. The shear is an anisotropic distortion, creating a major 'a' and minor 'b' axis which are equal to the new radius (R_0) scaled by the factors $1+\gamma$ and $1-\gamma$ respectively. (Source: Van Waerbeke & Mellier (2003))

Jacobian matrix (or amplification matrix), which describes how changes in position in the source plane (β) are related to changes in the observed position (θ):

$$\mathscr{A}(\theta) = \frac{\partial \beta}{\partial \theta} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{1.14}$$

where κ is identified as the convergence and $\gamma = \gamma_1 + i\gamma_2 = |\gamma| e^{2i\varphi}$ as the shear. The shape of a galaxy in the source plane gets distorted through this relationship and is observed to have a different shape. The shape distortion of galaxies by weak lensing can be quantified by the convergence, which is a uniform scaling of the galaxy image, and the shear, which is an anisotropic distortion. In the weak lensing regime both of these quantities are much smaller than unity, hence the amplification matrix is approximately equal to the identity matrix, and the image distortions are small.

Consider a circular source with radius R_0 (see Fig.1.2). In the absence of shear, ($\gamma = 0$), the image will be circular with a modified radius given by $R_0(1-\kappa)^{-1}$. Shearing will cause the image to be elliptical with $|\gamma| = (1-r)(1+r)^{-1}$, where $r = b/a$ is the ratio of the minor (b) to major (a) axis, and the direction of the shear is given by its phase φ. Note that in the simplified case of a circular source, the ellipticity ε of the image is a direct measure of the shear:

$$|\gamma| = |\varepsilon| = \frac{1-r}{1+r}. \tag{1.15}$$

In general, the intrinsic ellipticity of the source ε_s is non-zero, yielding (in the weak lensing limit, $\kappa << 1$ and $\gamma << 1$)

$$\varepsilon = \varepsilon_s + \gamma. \tag{1.16}$$

Although the intrinsic ellipticity of a given source is unknown, it is safe to assume that the average over source ellipticities will tend to zero; that is, there is no preferred direction for the intrinsic ellipticities.

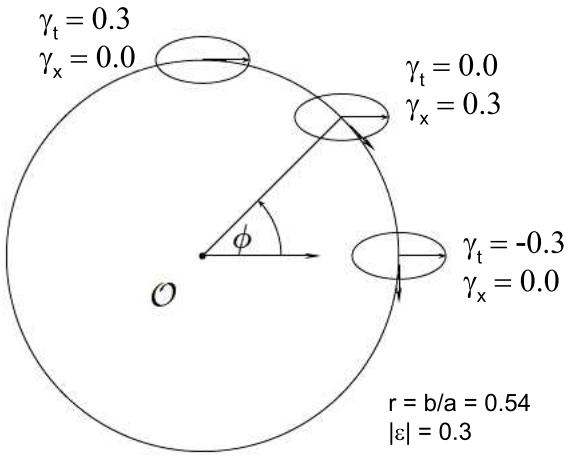

Figure 1.3: The shear decomposed into the tangential γ_t and cross-component γ_x. The sheared image has an ellipticity $|\varepsilon| = |\gamma| = 0.3$. (Source: Schneider (2005))

Hence, a sufficiently large sample of background galaxies will provide an accurate measure of the shear[4] $\langle\varepsilon\rangle = \langle\gamma\rangle$. This assumption can break down in certain regimes, which is discussed in Section 1.3.4.

The ellipticity is the primary observable in weak lensing analyses. Its interpretation as analogous to the shear will allow us to relate the measured shear correlation functions to the modelled matter density power spectrum. The shear γ is typically decomposed into two components: the tangential shear γ_t and the cross component γ_x, as shown in Figure 1.3. The shear components are defined relative to the separation vector for each pair of galaxies, with γ_t describing elongation and compression of the ellipticity along the separation vector and γ_x describing elongation and compression along a direction rotated 45° from the separation vector. The sign convention for γ_t is such that a galaxy sheared tangentially, as an arc around a central mass, has a positive γ_t. A negative γ_t describes a radially stretched ellipticity. The shear components are presented visually in Figure 1.3.

Pairs of galaxies are used because it is impossible to determine from individual galaxies whether lensing has occurred, since the weak lensing shape distortions are relative to the position of the lensing mass. We do not know where the lensing masses are on the sky, but we do know that they will induce characteristic shape distortions. Galaxies will tend to align their ellipticity tangentially to the center of the lensing mass. By considering the two-point correlation function, the relative alignment of an average pair of galaxies is quantified and can be compared to expectations from models. In the absence of weak lensing, one would expect there to be no preferential alignment and therefore no correlation of galaxy ellipticities except for intrinsic alignment, which is discussed in Section 1.3.4.

The two-point correlation functions are defined for each component $\langle\gamma_t\gamma_t\rangle(\theta)$ and $\langle\gamma_x\gamma_x\rangle(\theta)$, where all pairs separated by an angle θ are averaged. The following combination of these statistics is very useful since they can be measured directly from the data. Other second order statistics are often derived

[4]The practical details of accurately measuring the shapes of galaxies from data is beyond the scope of this work. The CFHTLenS data used in Chapter 3 has been analysed with a new Bayesian technique known as *lens*fit (Miller et al., 2013).

from these shear-shear correlation functions:

$$\begin{aligned}\xi_{\pm}(\theta) &= \langle\gamma_t\gamma_t\rangle(\theta) \pm \langle\gamma_x\gamma_x\rangle(\theta) \\ &= \frac{\Sigma_{i,j}\left[\gamma_{t,i}(\boldsymbol{\vartheta}_i)\gamma_{t,j}(\boldsymbol{\vartheta}_j) \pm \gamma_{x,i}(\boldsymbol{\vartheta}_i)\gamma_{x,j}(\boldsymbol{\vartheta}_j)\right]\Delta_{ij}}{N_p},\end{aligned} \tag{1.17}$$

where galaxy pairs labelled i and j are separated by angular distance $\vartheta = |\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_j|$. Here $\boldsymbol{\vartheta}_i$ is a two-dimensional vector given by the angular position of galaxy i. If ϑ falls in the angular bin given by θ, then Δ_{ij}=1; otherwise, Δ_{ij}=0. The number of pairs summed over is given by N_p. These correlation functions can be related to the convergence power spectrum, $P_\kappa(\ell)$ (e.g. Kaiser, 1992)

$$\begin{aligned}\xi_+(\theta) &= \frac{1}{2\pi}\int_0^\infty \mathrm{d}\ell\,\ell\,J_0(\ell\theta)P_\kappa(\ell) \\ \xi_-(\theta) &= \frac{1}{2\pi}\int_0^\infty \mathrm{d}\ell\,\ell\,J_4(\ell\theta)P_\kappa(\ell),\end{aligned} \tag{1.18}$$

where J_n represents the n^{th} order Bessel function of the first kind, and ℓ is the modulus of the two dimensional wave vector. The convergence power spectrum (defined mathematically below) quantifies on which scales weak lensing has the largest signal. In order to use the observable quantity in Equation (1.17) to constrain cosmology, we must find a cosmologically dependent model for the convergence power spectrum $P_\kappa(\ell)$. This is the goal of the rest of this subsection.

Defining mass density as ρ and the average mass density of the Universe as $\bar{\rho}$, the mass density contrast δ is

$$\delta(\theta) = \frac{\rho(\theta) - \bar{\rho}}{\bar{\rho}}. \tag{1.19}$$

The mass density power spectrum $P_\delta(k,\chi)$ is defined as

$$\langle\tilde{\delta}(\boldsymbol{k})\tilde{\delta}^*(\boldsymbol{k}')\rangle = (2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}')P_\delta(k,\chi), \tag{1.20}$$

where $\tilde{\delta}(\boldsymbol{k})$ is the Fourier transform of δ, the asterisk denotes complex conjugation, δ_D is the Dirac delta function, χ is the comoving distance, and k is the modulus of the three dimensional wave vector $\boldsymbol{k}$. The Dirac delta function arises due to the assumption of homogeneity and isotropy of the Universe on large scales. The mass density power spectrum describes the scales on which density fluctuations occur. If it is large for a particular value of k, then mass tends to cluster on the corresponding scales. Note that k is a reciprocal distance measure and large k correspond to small distances. The mass density power spectrum can be modelled from our understanding of the initial density fluctuations and how they evolve over time. The evolution is analytical in the linear regime of structure formation (where $\delta << 1$), but becomes highly non-linear at small scales and later times. Methods to account for non-linear structure growth have been devised by Peacock & Dodds (1996) and Smith et al. (2003a), following the work of Hamilton et al. (1991). Combining the linear and non-linear models provides a cosmology-dependent description of the shape of the power spectrum. The overall normalisation is a free parameter. The normalisation parameter is defined as the mass density variance within a sphere of 8 h^{-1}Mpc radius at

zero comoving distance (Peebles, 1993):

$$\sigma_8^2 = \langle \delta_R^2 \rangle = \frac{1}{2\pi^3} \int dk P_\delta(k, \chi = 0) |W(kR)|^2, \tag{1.21}$$

where $W(kR)$ is the Fourier transform of the top-hat window function. In real space, the top-hat window function is constant for $|x| < R$ and zero otherwise; its Fourier transform is

$$W(kR) = \frac{3}{(kR)^2} \left(\frac{\sin(kR)}{kR} - \cos(kR) \right). \tag{1.22}$$

Figure 1.4 shows how the mass density power spectrum changes with cosmological parameters and with different prescriptions for the non-linear growth of density perturbations.

The convergence power spectrum $P_\kappa(\ell)$ is defined to be:

$$\langle \tilde{\kappa}(\boldsymbol{\ell}) \tilde{\kappa}^*(\boldsymbol{\ell}') \rangle = (2\pi)^2 \delta_{\rm D}(\boldsymbol{\ell} - \boldsymbol{\ell}') P_\kappa(\ell), \tag{1.23}$$

where ℓ is the modulus of the two-dimensional wave vector perpendicular to the line of sight $\boldsymbol{\ell}$. We can express the convergence power spectrum in terms of a two-dimensional wave vector instead of a three-dimensional wave vector because the series of deflections a light bundle undergoes can be approximated by a single lensing plane. In the weak lensing regime, the correlation function of the shear in Fourier space is identical to the correlation function of the convergence:

$$\langle \tilde{\gamma}(\boldsymbol{\ell}) \tilde{\gamma}^*(\boldsymbol{\ell}') \rangle = \langle \tilde{\kappa}(\boldsymbol{\ell}) \tilde{\kappa}^*(\boldsymbol{\ell}') \rangle. \tag{1.24}$$

This equivalence is why the shear correlation functions in Equation (1.17) can be expressed in terms of the convergence power spectrum. The convergence power spectrum can also be written as

$$P_\kappa(\ell) = \frac{9 H_0^4 \Omega_{\rm m}^2}{4c^4} \int_0^{\chi_{\rm h}} {\rm d}\chi \frac{g^2(\chi)}{a^2(\chi)} P_\delta \left(k = \frac{\ell}{f_K(\chi)}, \chi \right), \tag{1.25}$$

where $\chi_{\rm h}$ is the comoving distance to the horizon[5], $a(\chi)$ is the scale factor, $f_K(\chi)$ is the comoving angular diameter distance (Equation 1.3), and $g(\chi)$ is the lensing efficiency which depends on the distribution of sources $n(\chi)$

$$g(\chi) = \int_\chi^{\chi_{\rm h}} {\rm d}\chi' \, n(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}. \tag{1.26}$$

It is evident from the above relationships that the strength of the weak lensing signal depends on cosmology. The convergence power spectrum can be modelled (Equation 1.25) for different cosmologies. A best fit cosmology can be determined through a standard likelihood analysis, with the likelihood given by

$$\mathscr{L} = \frac{1}{\sqrt{(2\pi)^{\rm n} |\mathbf{C}|}} \exp \left[-\frac{1}{2} (\mathbf{d} - \mathbf{m}) \mathbf{C}^{-1} (\mathbf{d} - \mathbf{m})^T \right], \tag{1.27}$$

[5]The furthest comoving distance from which light has had enough time to reach us.

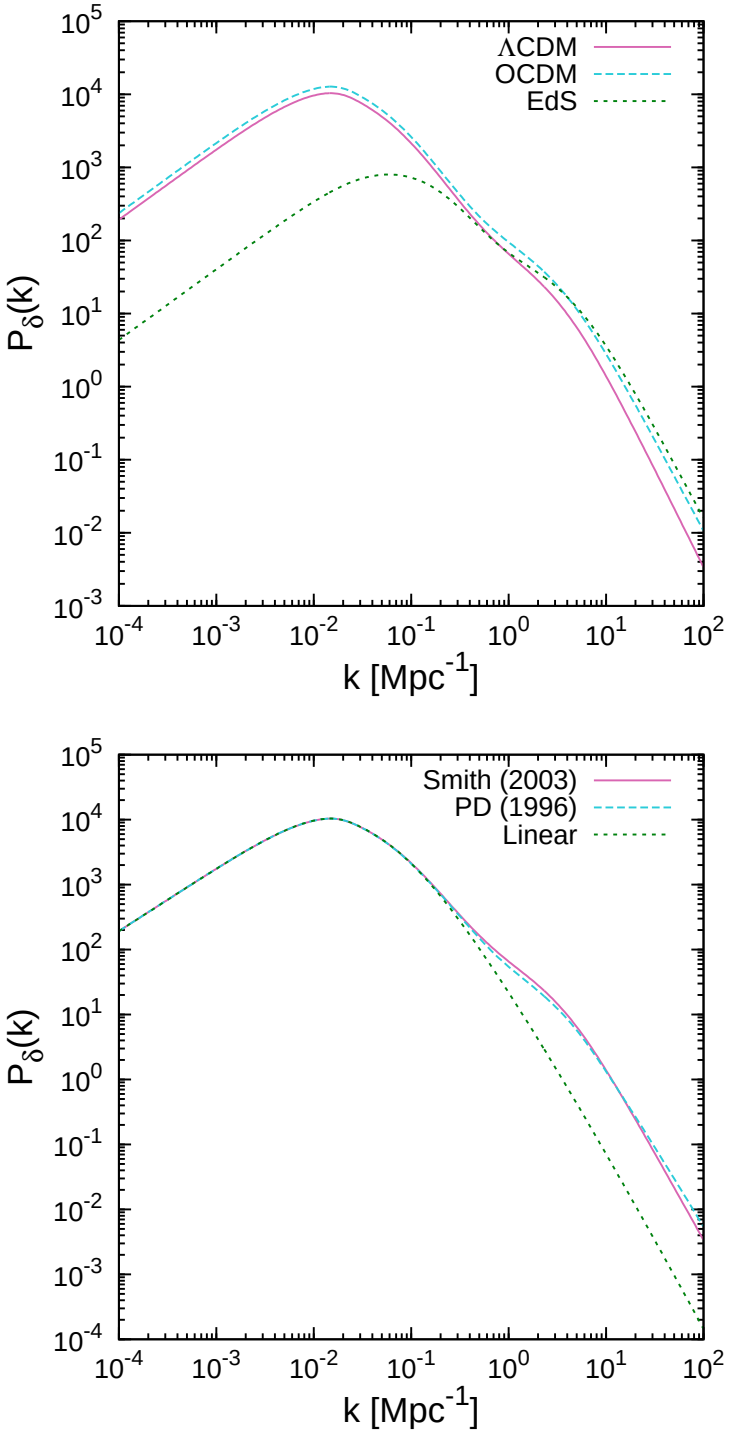

Figure 1.4: **Top:** How the mass density power spectrum (Equation 1.20) changes with various Cold Dark Matter (CDM) cosmologies. **ΛCDM (solid):** A flat cosmology ($\Omega_m + \Omega_\Lambda = 1$) with a non-zero cosmological parameter, $\Omega_m = 0.27$, and $\Omega_\Lambda = 0.73$. **OCDM (dashed):** An open cosmology ($\Omega_m + \Omega_\Lambda < 1$) with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.0$. **EdS (dotted):** An Einstein de-Sitter cosmology with $\Omega_m = 1.0$, $\Omega_\Lambda = 0.0$. For all cases $\sigma_8 = 0.78$ and the non-linear prescription of Smith et al. (2003a) is used. **Bottom:** The linear power spectrum is given by the dotted line, two methods are used to account for the non-linear power, Peacock & Dodds (1996) (dashed) and Smith et al. (2003a) (solid).

where **d** is the data vector $\xi_{\pm}(\theta)$ measured at n angular scales (Equation 1.17), **m** is the model vector for the same scales calculated from Equations (1.25) and (1.18), and **C** is the n $\times$ n covariance matrix. As seen from Eq.(1.25), weak lensing is most sensitive to a combination of the mass energy density at redshift zero $\Omega_{\rm m}$, the redshift distribution of the sources $n(\chi)$, the Hubble parameter at redshift zero H_0, and the normalisation of the mass density power spectrum σ_8. However, the most useful constraints are those placed on $\Omega_{\rm m}$ and σ_8 since they compliment constraints from other cosmological probes (see Chapter 3). Figure 1.5 shows how the convergence power spectrum changes with cosmology and non-linear modelling of the mass density power spectrum.

The above framework considers the gravitational lensing by LSS for one sample of source galaxies with a redshift distribution given by $n(\chi)$. In Chapter 3, we perform such an analysis on data from the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS), which is currently the largest weak lensing survey having an effective area on the sky of 154 square degrees. We extend this theoretical framework to include weak lensing from several redshift bins, known as weak lensing tomography. The main advantage of a tomographic analysis is to probe cosmological parameters as a function of redshift, which leads to better constrains on cosmological models. Breaking the sample into redshift bins requires well-measured redshifts. Unfortunately, spectroscopic redshifts for large and deep surveys are not practical, requiring the use of photometric redshifts. The dependence of weak lensing on photometric redshifts motivates the work presented in Chapter 2, which aims to better characterise photometric redshift distributions.

1.3.4 Intrinsic alignments

We saw in Equation (1.16) that the observed ellipticity of a galaxy (ε) depends on the source ellipticity (ε_s) and the shear (γ). The two point shear correlation function is measured by averaging over all galaxy pairs (i,j) separated by angle θ:

$$\begin{aligned} \langle \varepsilon^{\rm i} \varepsilon^{\rm j} \rangle &= \langle (\varepsilon_s^{\rm i} + \gamma^{\rm i})(\varepsilon_s^{\rm j} + \gamma^{\rm j}) \rangle, \\ &= \langle \varepsilon_s^{\rm i} \varepsilon_s^{\rm j} \rangle + \langle \varepsilon_s^{\rm i} \gamma^{\rm j} \rangle + \langle \varepsilon_s^{\rm j} \gamma^{\rm i} \rangle + \langle \gamma^{\rm i} \gamma^{\rm j} \rangle. \end{aligned} \tag{1.28}$$

Most weak lensing analyses can safely assume that the first three terms on the right hand side are sufficiently small to be ignored since they contain the source ellipticity term. Therefore, the two-point shear correlation function $\langle \gamma^{\rm i} \gamma^{\rm j} \rangle$ is equal to the correlation function of the observed galaxy ellipticities $\langle \varepsilon^{\rm i} \varepsilon^{\rm j} \rangle$. The three terms are known as intrinsic alignments since they all involve a correlation of the source ellipticity.

Early studies investigating the impact of intrinsic alignments focused on the so-called 'II' term resulting from a correlation of source ellipticities $\langle \varepsilon_s^{\rm i} \varepsilon_s^{\rm j} \rangle$. This term can be significantly non-zero if galaxies evolve in a common gravitational field that correlates their shapes. Therefore, the effect is most pronounced between physically nearby galaxies. Numerical studies (Croft & Metzler, 2000; Heavens et al., 2000), analytical studies (Catelan et al., 2001; Crittenden et al., 2001; Lee & Pen, 2001), and low redshift observational constraints (Brown et al., 2002a; Pen et al., 2000) of II were in close agreement.

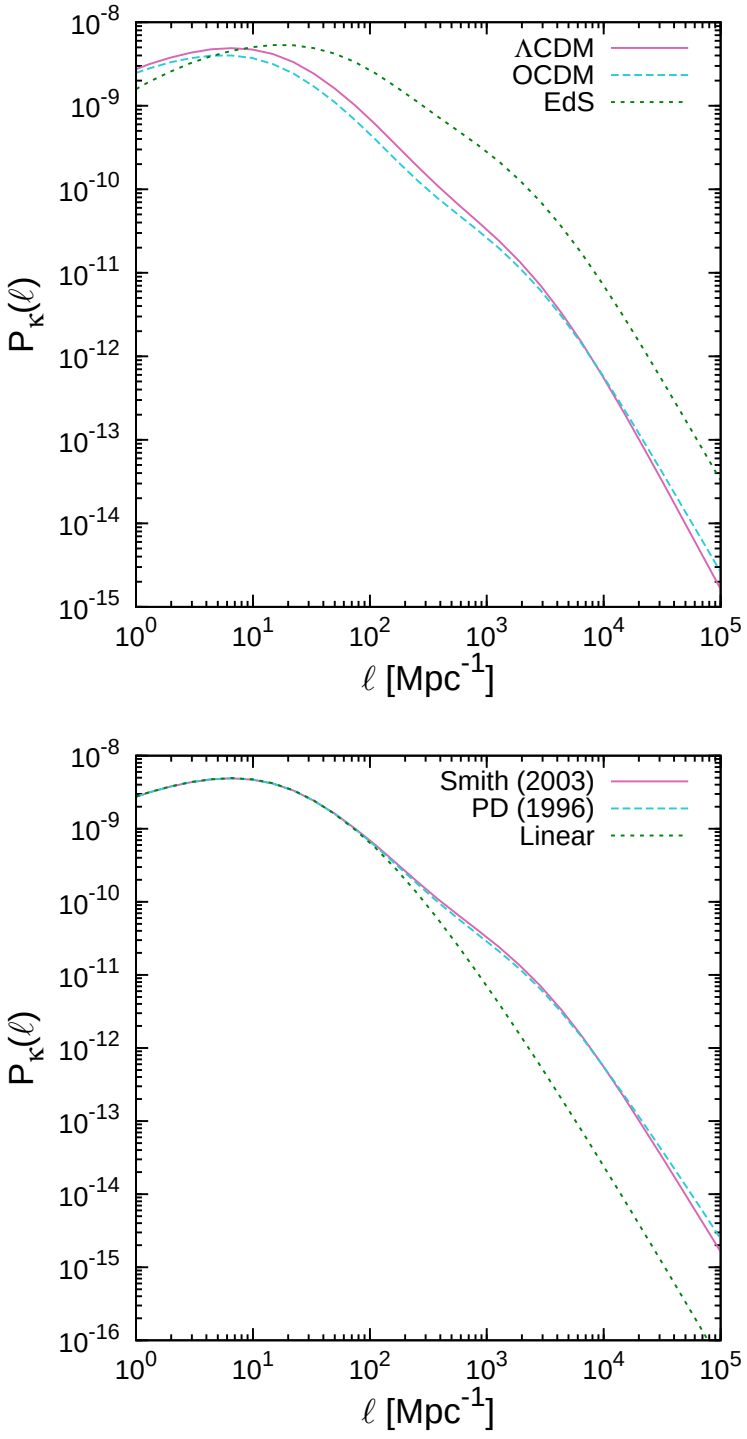

Figure 1.5: **Top:** How the convergence power spectrum (Equation 1.23) changes with various CDM cosmologies. **ΛCDM (solid):** $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$. **OCDM (dashed):** $\Omega_m = 0.27$, $\Omega_\Lambda = 0.0$. **EdS (dotted):** $\Omega_m = 1.0$, $\Omega_\Lambda = 0.0$. For all cases, $\sigma_8 = 0.78$ and the non-linear prescription of Smith et al. (2003a) is used. **Bottom:** The linear power spectrum is given by the dotted line, two methods are used to account for the non-linear power, Peacock & Dodds (1996) (dashed) and Smith et al. (2003a) (solid).

Deep weak lensing surveys could expect contamination of the measured shear correlation of at most a few per cent. However, King & Schneider (2003) and Heymans et al. (2004) were quick to point out that this effect could be much more significant if the galaxy sample was split into narrow redshift bins, which would enhance the relative number of close pairs. Heymans et al. (2004) also advocated for the down-weighting of close pairs to lessen the impact of the II term.

The second and third terms on the right hand side of Equation (1.28) are referred to as 'GI' since they involve the correlation of the intrinsic shape with the gravitational shear $\langle \varepsilon_s \gamma \rangle$. Hirata & Seljak (2004a) first demonstrated that the GI contribution can be significant. They showed that lensed source galaxies that get sheared by a foreground matter over-density may be correlated with foreground galaxies whose shapes are also correlated with the over-density. This effect cannot be eliminated by removing physically close pairs since the spurious correlation is between galaxies well separated in redshift.

Observations of low-redshift galaxies were used to constrain the magnitude of the II and GI terms, by cross-correlating the number density of galaxies (a proxy for the mass density) and the shape of galaxies. The first such study probed galaxies at $z\sim0.1$ in the Sloan Digital Sky Survey (SDSS) spectroscopic sample (Mandelbaum et al., 2006). The SDSS Luminous Red Galaxy sample (LRG, Hirata et al., 2007) and the MegaZ-LRG sample (Joachimi et al., 2011) extended this method to $z\sim0.6$. These studies discovered a dependence on galaxy type, with II and GI being more significant for red (late-type) galaxies, as opposed to blue (early-type). The atypical selection function of these works meant that the IA terms were not well characterised for a typical weak lensing survey, which contains far fewer LRGs. The results most applicable to deep weak lensing surveys come from Wiggle-Z (Mandelbaum et al., 2011) with a depth of $z\sim0.7$. Mandelbaum et al. (2011) report a null detection of the IA signal for blue galaxies but estimate that the error on the normalisation of the matter power spectrum (σ_8) for a CFHTLS-type survey[6] is ±0.03, well below the statistical error.

For tomographic analyses, intrinsic alignment contamination from both II and GI is more pronounced. This results from the splitting of galaxies into redshift bins, which increases the density of contaminating pairs compared to a 2D analysis. In the case of GI, the cross-correlation of distant redshift bins will enhance the GI signal, whereas these pairs would be diluted by including all redshift bins in a 2D analysis. Similarly, the auto-correlation of narrow redshift slices will enhance the II contamination, which would be diluted if a broad range of redshifts were included. In order to perform finely binned tomographic analyses, these IA contaminations needs to be addressed. King (2005) propose a model fitting technique to reduce the impact of IA contamination. They model the II and GI terms simultaneously with the cosmic shear signal and marginalise over the nuisance parameters introduced to model the intrinsic alignments. This method was extended by Bridle & King (2007a), who demonstrated that the predicted constraining power when marginalising over IAs was degraded with a severity depending on assumed prior knowledge of the intrinsic alignment signal, the accuracy of the photometric redshifts, and the flexibility of the IA model. This can be mitigated by simultaneously fitting galaxy cluster data which can calibrate the IA signal (Joachimi & Bridle, 2010; Zhang, 2010). This approach was first used by Kirk et al. (2010) combining lensing information from the 100 square degree lensing survey

[6] See Chapter 3, where the fully marginalised error on σ_8 for CFHTLenS is ±0.23

(Benjamin et al., 2007) and shear-shape clustering data from SDSS (Mandelbaum et al., 2006).

1.4 Photometric redshifts

Weak lensing surveys require knowledge of the radial distribution of galaxies (see equations 1.25 and 1.26). The statistical nature of the lensing signal requires large surveys with large galaxy number densities, leading to millions of detected galaxies. It is not possible to measure spectroscopic redshifts ($z_{\rm spec}$) for all galaxies in weak lensing surveys since faint galaxies require long periods of observation to secure reliable spectroscopic redshifts. The alternative is the use of photometric redshifts (z_p), which are far less accurate but are the only tool capable of estimating the redshifts of millions of galaxies in a reasonable amount of time.

Spectroscopic redshifts are obtained by diffraction of light from a galaxy. This allows the intensity of light to be measured as a function of wavelength, and fine details such as individual absorption and emission lines can be detected. The amount of light energy per unit wavelength is referred to as the spectral energy distribution (SED). Figure 1.6 shows examples of SEDs for several types of galaxies. If individual features in an SED can be identified, then the redshift can be determined by the observed wavelength (λ_o) of these features,

$$z = \frac{\lambda_o - \lambda_e}{\lambda_e}, \tag{1.29}$$

where λ_e is the wavelength of emission for this feature. This is nothing more than the familiar Doppler Effect, which causes the observed wavelength of light to increase if the source emitting that light is moving away from the observer.[7]

Measuring the SEDs of galaxies is time consuming since the energy in small intervals of wavelength is measured. This becomes worse for faint galaxies. For this reason, large surveys measure the flux of galaxies in large intervals of wavelength. These intervals are called broadband filters or photometric bands. A common set of filters and the one used in Chapters 2 and 3 is presented in the bottom right panel of Figure 1.6. The bands are labelled u^*, g', r', i', and z'. The transmission curves in Figure 1.6 quantify what percentage of light energy will reach the detector for a given filter. The transmission curves are combined with the quantum efficiency of the detector to fully describe the fraction of incident light that is detected as a function of wavelength. Observing a galaxy in these broadband filters takes far less time than obtaining a detailed spectra.

Photometric redshifts can be estimated by comparing the observed photometry of a galaxy against a template SED that has been convolved with the filter transmission curves. The template spectrum can be shifted in wavelength to simulate a higher redshift galaxy. Model fitting techniques are used to estimate the best-fitting template. The result is an estimate of the redshift of the galaxy. Typically, spectroscopic redshifts are used either as a training set to iteratively improve the model fitting, or as a test sample with which to quantify errors in the derived photometric redshifts. The spectroscopic sample will usually only cover a small fraction of the entire survey area and have a much brighter magnitude limit. An example of such a comparison is presented in Figure 1.7. Note that these are the redshifts used

[7] Note that the redshift and scale factor introduced in Section 1.1 are closely related: $a(t) = (1+z)^{-1}$.

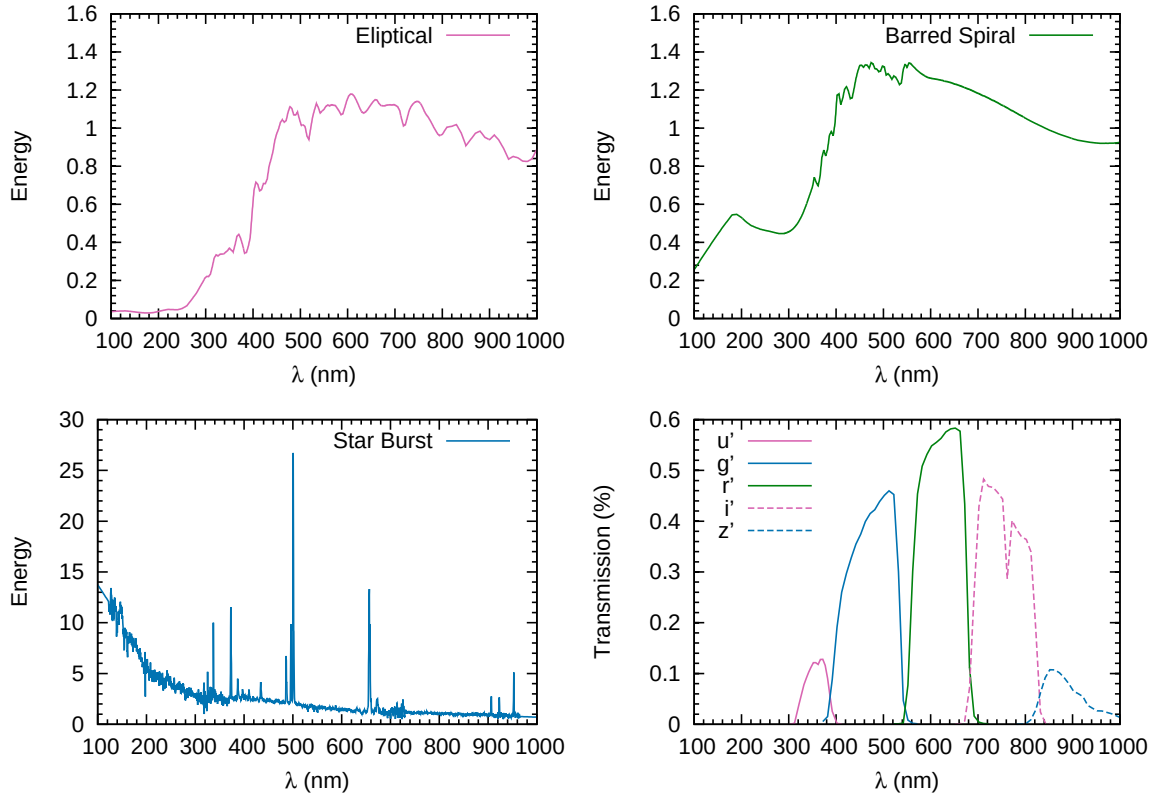

Figure 1.6: Three template spectral energy distributions are shown for different galaxy types. These templates are taken from Benítez et al. (2004). Each template shows the energy distribution as a function of wavelength (λ) for a galaxy of the indicated type at redshift zero (i.e. the wavelength of emission). The panel in the bottom right shows the transmission curves of five broad band filters. The convolution of the broad band filters with an SED, which can be shifted to higher wavelengths to simulate a higher redshift galaxy, yields the observed colours of the galaxy.

in Chapter 3.

In this work, we will discuss only the template fitting method (described above) for obtaining photometric redshifts; however, several other methods exist. For an overview of photometric redshift methods and codes used in the literature, the reader is referred to the PHoto-z Accuracy Testing programme (PHAT, Hildebrandt et al., 2010). PHAT was undertaken in order to quantify the performance of the many different photometric redshift methods and codes that are used today. For purposes of this thesis, it is important to remember that regardless of the method used, all photometric redshifts contain errors. The type of scatter seen in Figure 1.7 is generic to all photometric redshift methods.

The scatter in Figure 1.7 is due primarily to uncertainties in the observed photometry and subsequent imperfect model fitting. A galaxy may not be well described by the template SEDs, or there may be a degeneracy between two or more template SEDs at different redshifts. The scatter is often described by the standard deviation in the quantity

$$\Delta z = \frac{z_{\rm phot} - z_{\rm spec}}{1 + z_{\rm spec}}, \tag{1.30}$$

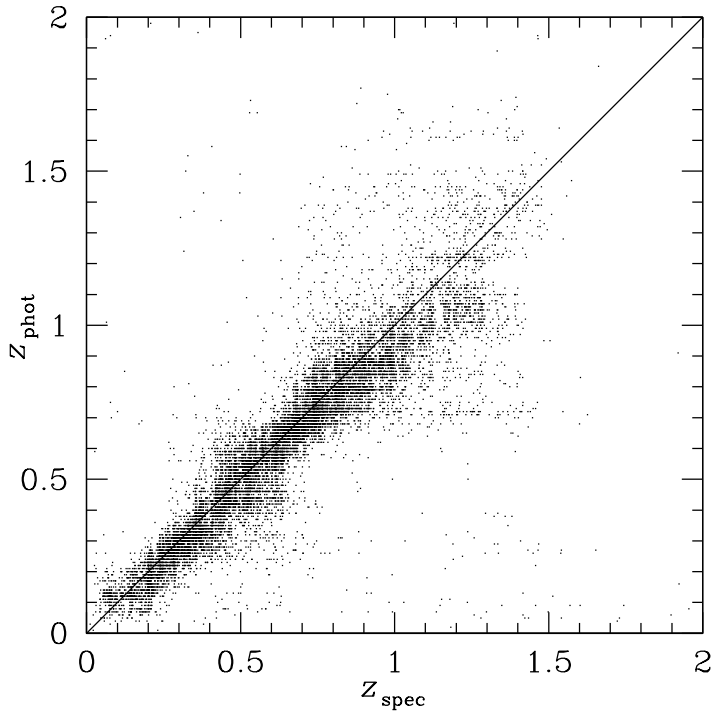

Figure 1.7: Comparison of the photometric redshifts ($z_{\rm phot}$) to the spectroscopic redshifts ($z_{\rm spec}$) for CFHTLenS (Hildebrandt et al., 2012). If the photometric redshifts were perfect, all the points would fall on the plotted $z_{\rm phot} = z_{\rm spec}$ line.

denoted by $\sigma_{\Delta z}$. The scatter in Figure 1.7 is noticeably worse at very low redshift $z_{\rm phot} \lesssim 0.1$ and very high redshift $z_{\rm spec} \gtrsim 1$. A detailed analysis in Hildebrandt et al. (2012) determines a high confidence photometric redshift range of $0.1 \leq z_{\rm phot} < 1.3$ for which the scatter is $0.03 < \sigma_{\Delta z} < 0.06$.

There are several outliers in Figure 1.7; for example, there is a group of galaxies predicted to be at $z_{\rm phot} < 0.5$ but with $z_{\rm spec} > 1$. In Chapter 2, we present a method to measure contamination in the photometric redshifts well suited to detecting this type of catastrophic error. This is important to weak lensing analyses since low redshift galaxies are not strongly lensed. If there are high redshift galaxies contaminating the low redshift sample, the observed lensing signal will be higher than expected. Therefore, the derived cosmological parameters will be biased. Understanding the scatter between photometric redshifts in order to improve the accuracy of cosmological parameter constraints from weak lensing is one of the central topics of this thesis.

In Chapter 3, we investigate the accuracy of the posterior probability distribution function (PDF) of the photometric redshifts. For each galaxy, a PDF can be estimated, which quantifies the probability of finding the galaxy as a function of redshift (see Figure 1.8). This is achieved by using a Bayesian photometric redshift technique described by Benítez (2000) and applied to the CFHTLenS data set by Hildebrandt et al. (2012). We show several representative PDFs in Figure 1.8. The legend displays the best-fitting redshift, which is taken as the peak of the PDF. Note that some galaxies have significant

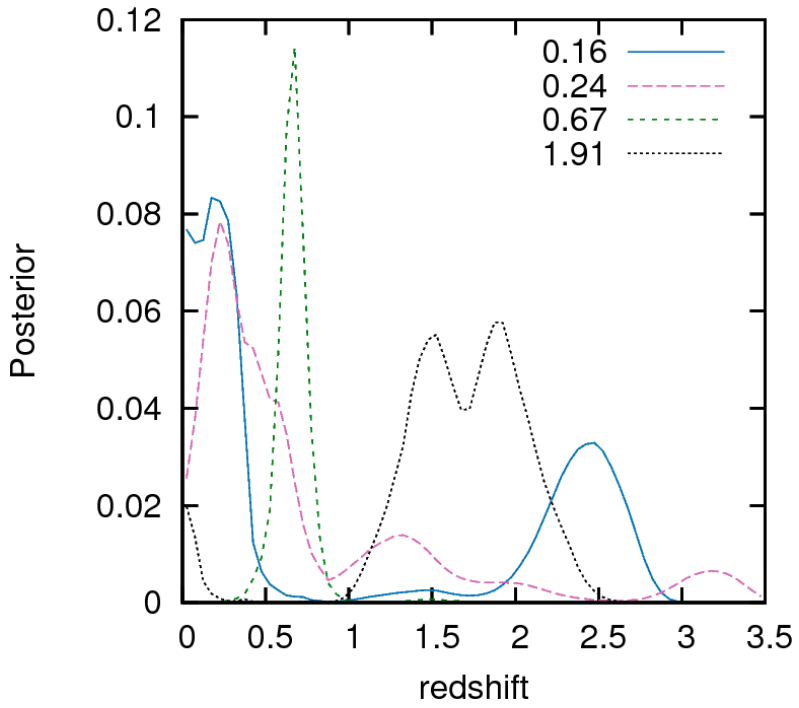

Figure 1.8: The posterior probability distribution functions (PDF) for several characteristic galaxies selected from the CFHTLenS catalogue (Hildebrandt et al., 2012). The label provides the best-fitting redshift estimate, which is taken to be the peak of the PDF.

degeneracies between high and low redshifts. In Chapter 3, we investigate the accuracy of using the entire PDF, showing that this provides a better estimate of the redshift distribution of the galaxies.

1.5 Thesis overview

This thesis is composed of two related research projects. The first, presented in Chapter 2, derives a novel method to measure contamination between photometric redshift bins by employing the angular correlation function. The theoretical details of the method are derived and it is tested on both simulated data and survey data. The importance of accurate photometric redshifts for weak lensing cosmology is highlighted in Chapter 2 and, in fact, provided the motivation for the research project. The second project is presented in Chapter 3, and comprises two scientific objectives:

1. Use the method presented in Chapter 2 to analyse the photometric redshifts of the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) and determine if the photometric redshifts' Probability Distribution Functions (PDF) accurately account for redshift contamination.

2. Use two redshift bins to perform a tomographic weak lensing analysis of CFHTLenS in order to obtain cosmological parameter constraints and verify that the redshift scaling of the shear signal agrees with expectations.

We conclude and briefly discuss future work in Chapter 4.

Chapter 2

Photometric redshifts: estimating their contamination and distribution using clustering information[1]

2.1 Introduction

The time-intensive nature of spectroscopy and the immense number of galaxies in current and future surveys have secured the place of photometric redshifts among the most valuable astronomical tools. Nearly all cosmological probes are sensitive to distance, making redshifts particularly important to cosmology. While spectroscopy is the most accurate and precise way to measure redshift, it is far too time-intensive to apply to current or future surveys which require redshift estimates for millions of galaxies. Hence, photometric redshifts and a thorough understanding of their uncertainties are vital to the study of large galaxy surveys.

Degeneracies in fitting multi-band photometric data can result in significant misidentification causing contamination between all redshifts (see Section 1.4). In this chapter we investigate contamination between photometric redshift bins and present a method for estimating the contamination using the angular correlation function. This method relies only on the photometric redshifts; it does not require a spectroscopic sample. However, it is likely that the photometric redshifts will have been calibrated with a spectroscopic training sample.

Quantifying photometric redshift contamination is essential to exploit the full potential of future photometric surveys, such as the Large Synoptic Survey Telescope (LSST) or the Supernova Acceleration Probe (SNAP), which aim to constrain dark energy. Weak lensing and baryon acoustic oscillations (BAO) are both sensitive to the mean redshift of a bin, requiring an unbiased measurement-accuracy on the order of $0.001 - 0.005$ in redshift so that the constraints on dark energy are not significantly

[1] A version of this chapter has been published. J. Benjamin, Ludovic Van Waerbeke, Brice Ménard, and Martin Kilbinger. Photometric redshifts: estimating their contamination and distribution using clustering information, 2010, MNRAS, Volume 408, Issue 2, pp. 1168-1180

degraded (Huterer et al., 2004).

Several studies have investigated the use of the angular correlation function in determining the true redshift distribution of galaxies binned by photometric redshift (Newman, 2008; Schneider et al., 2006; Zhang et al., 2009). These works have adopted LSST-like survey parameters and focused on the Fisher information matrix as a means of forecasting constraints. Schneider et al. (2006) show that the angular correlation function in the linear regime can be used to measure the mean redshift to an accuracy of 0.01, noting that there is further constraining power at smaller scales. The method investigated by Newman (2008) relies on an overlapping spectroscopic sample so that the angular cross-correlation between the spectroscopic and the photometric redshifts can be exploited. They find that the desired accuracy on the mean redshift can be reached provided a spectroscopic sample of 25,000 galaxies per unit redshift. Zhang et al. (2009) investigate a self-calibration method, where the mean redshift in a bin is estimated from angular cross-correlations between photometric redshift bins, very similar to the one presented in this work and Erben et al. (2009). They demonstrate that self-calibration can reach the required accuracy on the mean redshift if additional information from weak lensing shear is used to help break parameter degeneracies.

We focus here on the practical application of measuring photometric contamination in both simulated and real data without the use of spectroscopic redshifts. This is of interest not only for meeting the stringent requirements of future surveys as mentioned above, but also for other applications such as: a diagnostic tool for photometric redshifts; determining background samples for cluster lensing and estimating the true redshift distribution for 2-D cosmic shear measurements. The details of our method for a strict two-bin analysis of the CFHTLS-Archive-Research-Survey (CARS) are presented in Erben et al. (2009), where we demonstrate that the contamination present in the bright spectroscopic training sample is consistent with the contamination seen in the much deeper photometric redshift sample. This addresses a central concern when calibrating photometric redshifts with spectroscopic redshifts.

This chapter is organised as follows: The angular correlation function and the estimators we use are presented in Section 2.2. The details of the analytic method for estimating redshift bin contamination are discussed in Section 2.3, which first addresses the general problem of contamination between an arbitrary number of redshift bins before focusing on the two-bin case and its extension to multiple redshift bins. The analytic model is tested using mock observational catalogues in Section 2.4, where we show that contamination between redshift bins can be accurately determined with our model. In Section 2.5 we measure contamination in a real galaxy survey demonstrating the ability of the method to constrain the true (uncontaminated) redshift distribution and measure the average redshift of each photometric redshift bin. Section 2.6 presents details of the maximum likelihood method, including a discussion of the covariance matrix. The contamination model for three redshift bins is discussed in Section 2.7. Concluding remarks, including limitations of the method and ideas for future work, are presented in Section 2.8.

2.2 Angular correlation function and estimators

The two-point angular correlation function describes the amount of clustering in a distribution of galaxies relative to what would be expected from a random distribution. The angular correlation function ω can be interpreted as the excess probability of finding an object in the solid angle dΩ a distance θ from another object. The total probability is given by

$$dP = N[1 + \omega(\theta)]\mathrm{d}\Omega, \tag{2.1}$$

where N is the density of objects per unit solid angle (Peebles, 1980). Note that $N\mathrm{d}\Omega$ is the Poisson probability that the solid angle element is occupied by a galaxy.

Many estimators have been devised for the correlation function. Kerscher et al. (2000) present a comparison of the most widely used estimators. The estimators primarily differ in the handling of edge effects, and for arbitrarily large number densities they all converge. In Section 2.3 we employ the simplest estimator,

$$\omega_{\mathrm{ij}} = \frac{(D_{\mathrm{i}}D_{\mathrm{j}})_\theta}{(RR)_\theta}\frac{N_{\mathrm{R}}N_{\mathrm{R}}}{N_{\mathrm{i}}N_{\mathrm{j}}} - 1, \tag{2.2}$$

where $(D_{\mathrm{i}}D_{\mathrm{j}})_\theta$ is the number of pairs separated by a distance θ between data sets i and j, $(RR)_\theta$ is the number of pairs separated by a distance θ for a random set of points, N_{R} is the number of points in the random sample and N_{i} (N_{j}) is the number of points in data sample i (j). In this work we consider i and j to be non-overlapping redshift bins. The auto-correlation is described by the case i=j, and the cross-correlation by the case i$\neq$j.

A more robust estimator (Landy & Szalay, 1993) is

$$\omega_{\mathrm{ij}} = \frac{(D_{\mathrm{i}}D_{\mathrm{j}})_\theta}{(RR)_\theta}\frac{N_{\mathrm{R}}N_{\mathrm{R}}}{N_{\mathrm{i}}N_{\mathrm{j}}} - \frac{(D_{\mathrm{i}}R)_\theta}{(RR)_\theta}\frac{N_{\mathrm{R}}}{N_{\mathrm{i}}} - \frac{(D_{\mathrm{j}}R)_\theta}{(RR)_\theta}\frac{N_{\mathrm{R}}}{N_{\mathrm{j}}} + 1. \tag{2.3}$$

This is used when measuring the correlation function from data, either the Millennium Simulation in Section 2.4 or the CFHTLS-Deep fields in Section 2.5.

Galaxies cluster in over-dense regions, leading to an excess number of pairs when compared to a random distribution of points. On small scales, the angular auto-correlation function of galaxies is positive for all redshifts, though the shape and amplitude vary as a function of redshift due to the evolution of structure formation. The angular cross-correlation between two distant redshift bins should be zero since galaxies that are physically separated by large distances do not cluster with one another.

When considering galaxies binned in redshift, neighbouring bins may have a significant cross-correlation, especially if the bin width is narrow, since a significant number of galaxies in each bin could be clustered with each other. In the case of photometric redshifts the typical redshift error ($\Delta z \approx 0.05$) will result in a large cross-correlation if the width of the bins is not much larger than this error.

Photometric redshift bins that are not neighbouring should not be physically clustered with one another and their cross-correlation should be zero. Deviations from zero indicate that the two bins contain galaxies that are physically clustered. These galaxies may result from contamination between

the two bins or from the mutual contamination of both bins from other redshifts.

Weak lensing magnification can cause galaxies at high redshift to cluster near lower redshift galaxies. This effect can be calculated and accounted for –we discuss this in Section 2.8.

2.3 Analytic development of the contamination model

The goal of the contamination model is to measure the level of contamination between all photometric redshift bins by measuring the angular cross-correlation. We have presented the details of the two-bin model in Erben et al. (2009) and applied it to the CFHTLS-Archive-Research Survey (CARS). Here we develop a fully consistent multi-bin approach in Section 2.3.1 before revisiting the two-bin case and extending it to a global pairwise analysis in Section 2.3.2.

2.3.1 Multi-bin analysis

We first address the case of an arbitrary number of redshift bins, where each bin is potentially contaminating every other bin. The number of galaxies observed to be in the i$^{\rm th}$ bin is $N_{\rm i}^{\rm o}$ which, by virtue of the mixing between bins, is not equal to the true number of galaxies in that redshift bin $N_{\rm i}^{\rm T}$. We define $f_{\rm ij}$ to be the number of galaxies from bin i contaminating bin j as a fraction of the true number of galaxies in bin i. Therefore $f_{\rm ij}N_{\rm i}^{\rm T}$ is the number of galaxies from bin i present in bin j. The fraction of galaxies in bin i which do not contaminate other bins is taken to be $f_{\rm ii}$ which is convenient shorthand for

$$f_{\rm ii} = 1 - \sum_{\rm k\neq i}^{\rm n} f_{\rm ik}, \tag{2.4}$$

where n is the number of redshift bins. The summation convention here, and throughout, assumes the sum is begins at one.

The observed number of galaxies in bin i will be those galaxies which contaminate the bin, plus those galaxies from bin i which do not contaminate other bins,

$$N_{\rm i}^{\rm o} = \sum_{\rm k}^{\rm n} N_{\rm k}^{\rm T} f_{\rm ki}, \tag{2.5}$$

where the k=i term accounts for those galaxies that remain in bin i while all other terms account for galaxies that have contaminated bin i. This results in a system of n equations which can be written as:

$$\begin{pmatrix} N_1^{\rm o} \\ N_2^{\rm o} \\ .. \\ N_{\rm n}^{\rm o} \end{pmatrix} = \begin{pmatrix} f_{11} & f_{21} & \dots & f_{\rm n1} \\ f_{12} & f_{22} & \dots & f_{\rm n2} \\ \dots & \dots & \dots & \dots \\ f_{1\rm n} & f_{2\rm n} & \dots & f_{\rm nn} \end{pmatrix} \begin{pmatrix} N_1^{\rm T} \\ N_2^{\rm T} \\ \dots \\ N_{\rm n}^{\rm T} \end{pmatrix}. \tag{2.6}$$

If the n $\times$ n matrix is invertible, then we can solve the system for the true number of galaxies in each bin. In the limit of zero contamination the off-diagonal elements will tend to zero while the diagonal elements tend to unity. In this limit the matrix is trivially non-singular. If a matrix is strictly diagonally

dominant then it follows from the Gershgorin circle theorem that it is non-singular. Therefore as long as

$$f_{\rm ii} > \sum_{\rm k\neq i}^{\rm n} f_{\rm ik}, \tag{2.7}$$

the matrix is strictly diagonally dominant and therefore is invertible. This condition is simply stating that a solution exits if the majority of the galaxies from the i$^{\rm th}$ true redshift bin do not contaminate other bins. The case of uniform contamination, where each bin sends $N_{\rm i}^{\rm T}/{\rm n}$ galaxies to each other bin, results in a singular matrix where all rows are identical.

The observed correlation functions can be derived by investigating which pairs contribute when correlating bin i and bin j. The observed number of data pairs between bins can be related to contributions from the true number of pairs,

$$(D_{\rm i}D_{\rm j})^{\rm o}_\theta = \sum_{\rm k}^{\rm n}\sum_{\rm l}^{\rm n}(D_{\rm k}D_{\rm l})^{\rm T}_\theta\, f_{\rm ki}\, f_{\rm lj}. \tag{2.8}$$

Multiplying both sides by $\frac{1}{(RR)_\theta}\frac{N_{\rm R}N_{\rm R}}{N_{\rm i}N_{\rm j}}$ and using Equation (2.2) to relate the pair counts to the correlation function yields

$$\omega^{\rm o}_{\rm ij} = \sum_{\rm k}^{\rm n}\sum_{\rm l}^{\rm n}\omega^{\rm T}_{\rm kl}\frac{N^{\rm T}_{\rm k}N^{\rm T}_{\rm l}}{N^{\rm o}_{\rm i}N^{\rm o}_{\rm j}}\, f_{\rm ki}\, f_{\rm lj}, \tag{2.9}$$

which can also be derived by considering $\left\langle N^{\rm o}_{\rm i}, N^{\rm o}_{\rm j}\right\rangle$.

We assume that the true cross-correlation between any two redshift bins is zero. Hence, it is useful to rewrite Equation (2.9) as

$$\omega^{\rm o}_{\rm ij} = \sum_{\rm k}^{\rm n}\omega^{\rm T}_{\rm kk}\frac{(N^{\rm T}_{\rm k})^2}{N^{\rm o}_{\rm i}N^{\rm o}_{\rm j}}\, f_{\rm ki}\, f_{\rm kj}. \tag{2.10}$$

Note that this allows us to express the observed auto- and cross-correlations as linear combinations of the true auto-correlation functions. The following matrix equation follows from the above when we consider only the equations for the observed auto-correlations (i.e, when i=j):

$$\begin{pmatrix} \omega^{\rm o}_{11}N^{\rm o^2}_1 \\ \omega^{\rm o}_{22}N^{\rm o^2}_2 \\ \cdots\cdots \\ \omega^{\rm o}_{\rm nn}N^{\rm o^2}_{\rm n} \end{pmatrix} = \begin{pmatrix} f^2_{11} & f^2_{21} & \cdots & f^2_{\rm n1} \\ f^2_{12} & f^2_{22} & \cdots & f^2_{\rm n2} \\ \cdots & \cdots & \cdots & \cdots \\ f^2_{1\rm n} & f^2_{2\rm n} & \cdots & f^2_{\rm nn} \end{pmatrix} \begin{pmatrix} \omega^{\rm T}_{11}N^{\rm T^2}_1 \\ \omega^{\rm T}_{22}N^{\rm T^2}_2 \\ \cdots\cdots \\ \omega^{\rm T}_{\rm nn}N^{\rm T^2}_{\rm n} \end{pmatrix}. \tag{2.11}$$

We are again left with an n $\times$ n matrix which, when inverted, lets us express the true auto-correlations in terms of the observed auto-correlations. Once they are known, we can use Equation (2.10) to express the observed cross-correlations as a linear combination of observed auto-correlation functions,

$$\omega^{\rm o}_{\rm ij} = \sum_{\rm k}^{\rm n}\omega^{\rm o}_{\rm kk}\frac{(N^{\rm o}_{\rm k})^2}{N^{\rm o}_{\rm i}N^{\rm o}_{\rm j}}g_{\rm k}(f), \tag{2.12}$$

where i $\neq$ j and $g_{\rm k}(f)$ is a complicated function of the contamination fractions. Note that the true number of objects from Equation (2.10) has cancelled with that from Equation (2.11).

The n $\times$ n matrix in Equation (2.11) above is very similar to that found in Equation (2.6). The

diagonal elements tend to unity as the contamination between bins tends to zero, resulting in a nearly diagonal matrix. The matrix will be strictly diagonally dominant and hence non-singular under the same condition found above, and for uniform contamination the matrix is singular.

The total number of unknown parameters (f_{ij}, where $i \neq j$) is $n(n-1)$. This is clear since, in general, each of the n bins can contaminate each of the other bins $(n-1)$. Our goal is to constrain these parameters with Equation (2.12), which yields only $\frac{n(n-1)}{2}$ equations since $\omega^o_{ij} = \omega^o_{ji}$. If we could only measure the angular correlation function at one scale then it would be impossible to constrain the parameters. However, if we have two independent measurements of the cross-correlations at different scales, we double the number of equations and are able to constrain the problem. In practice the measurements of the angular correlation function at different scales are not independent, and we should strive to include as large a range of scales as possible.

Equation (2.12) relates the amplitude of the cross-correlation to a weighted sum of the auto-correlations. If the shape of all the auto-correlations were the same, then it would be impossible to distinguish their contributions to the cross-correlation, which would also have the same shape. For example if the auto-correlations were well described by power laws with the same slope. In such cases this method would yield completely degenerate parameter constraints.

2.3.2 Pairwise analysis

Here we first consider the case of exactly two redshift bins. This provides a simple case with which to test our method using the Millennium Simulation in Section 2.4. We then expand this to multiple redshift bins by considering each pair in turn. We show that this global pairwise analysis will yield accurate estimates for the entire set of contamination fractions if the contamination is sufficiently small. The case of exactly three redshift bins is presented in Section 2.7.

We consider two redshift bins, labelled 1 and 2. From Equation (2.6), we can express the observed number of galaxies in each bin in terms of the true number of galaxies and the contamination fractions,

$$\begin{aligned} N^o_1 &= N^T_1(1-f_{12}) + N^T_2 f_{21}, \\ N^o_2 &= N^T_2(1-f_{21}) + N^T_1 f_{12}. \end{aligned} \tag{2.13}$$

Each observed redshift bin contains those galaxies which do not contaminate the other bin (e.g. $N^T_1(1-f_{12})$) plus those galaxies which are contaminating from the other bin (e.g. $N^T_2 f_{21}$). Note that the total number of galaxies, $N^o_1 + N^o_2 = N^T_1 + N^T_2$, is conserved.

Inverting Equation (2.11) allows us to express the true auto-correlations in terms of the observed auto-correlations. The resulting equalities can be plugged into Equation (2.10) yielding

$$\omega^o_{12} = \frac{\omega^o_{11}\frac{N^o_1}{N^o_2} f_{12}(1-f_{21}) + \omega^o_{22}\frac{N^o_2}{N^o_1} f_{21}(1-f_{12})}{((1-f_{12})(1-f_{21}) + f_{12}f_{21})}. \tag{2.14}$$

As long as the two bins considered comprise the entire sample of galaxies, this formalism is consistent. Considering a sub-sample allows for leakage to and from the region exterior to the two bins.

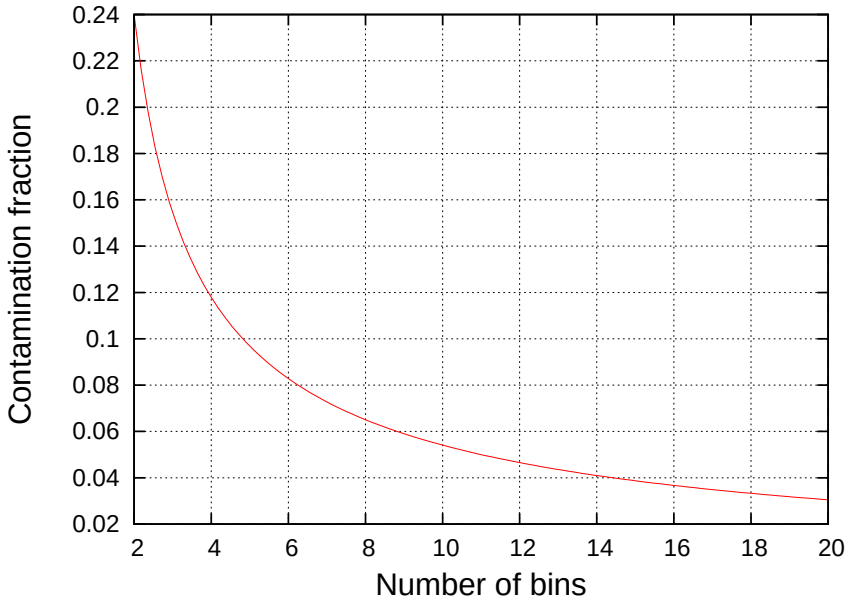

Figure 2.1: Upper bounds on the contamination fraction, f, as a function of the number of redshift bins n. The upper bound represents the limit at which the pairwise approach breaks down. A global contamination at these levels will cause at most a 10 per cent error in the estimate of ω_{ii}^{T} (Equation 2.15).

This can induce a cross-correlation due to mutual contamination of the considered bins from the exterior region and breaks the implicit assumption of galaxy number conservation.

The two-bin case has practical applications such as background selection in cluster lensing, where one seeks a background population that does not share members with the selected foreground cluster.

We now address the global pairwise analysis which –if the off diagonal contamination fractions are small– is a good approximation of the full-matrix approach detailed at the beginning of this section. In this case considering each pair of redshift bins in turn will yield estimates for the entire set of contamination fractions. We start by assuming that second order terms in the cross-contaminations are small, such that the off-diagonal elements of the matrix in Equation (2.11) satisfy $f_{ij}^2 \ll f_{ii}^2$. This results in the following simple relationship between the observed and true auto-correlations:

$$\omega_{ii}^{T} = \omega_{ii}^{o} \left(\frac{N_i^o}{N_i^T} \right)^2 \frac{1}{f_{ii}^2}. \tag{2.15}$$

When combined with Equation (2.10) this yields the following equation for the observed cross-correlation:

$$\omega_{ij}^{o} = \omega_{ii}^{o} \frac{N_i^o}{N_j^o} \frac{f_{ij}}{f_{ii}} + \omega_{jj}^{o} \frac{N_j^o}{N_i^o} \frac{f_{ji}}{f_{jj}} + \sum_{k,k\neq i,k\neq j}^{n} \omega_{kk}^{o} \frac{(N_k^o)^2}{N_i^o N_j^o} \frac{f_{ki} f_{kj}}{(f_{kk})^2}, \tag{2.16}$$

All terms that are first order in f have been taken out of the summation operator. This equation is identical to the two bin case of Equation (2.14) if we assume that second order terms are negligible.

This approximation will hold as long as the eigenvalues of the n×n matrix of Equation (2.11) are well approximated by the diagonal elements. To attempt to understand the limitation of this approximation we assume the severe case of uniform cross contamination. In this case all $f_{\rm ij}$ equal f, therefore $f_{\rm ii} = (1-({\rm n}-1)f)$ yielding the following contamination matrix:

$$\begin{pmatrix} (1-({\rm n}-1)f)^2 & f^2 & \dots & f^2 \\ f^2 & (1-({\rm n}-1)f)^2 & \dots & f^2 \\ \dots & \dots & \dots & \dots \\ f^2 & f^2 & \dots & (1-({\rm n}-1)f)^2 \end{pmatrix}. \tag{2.17}$$

The eigenvalues are $(1-({\rm n}-1)f)^2 - f^2$ with multiplicity ${\rm n}-1$ and $(1-({\rm n}-1)f)^2 + ({\rm n}-1)f^2$. The latter eigenvalue, being the larger deviation from the case of a purely diagonal matrix, gives us a constraint on f. If we require that the deviation be at most 10 per cent, we find

$$\frac{({\rm n}-1)f^2}{(1-({\rm n}-1)f)^2} = 0.1, \tag{2.18}$$

$$f = \begin{cases} \frac{1-(0.1({\rm n}-1))^{-0.5}}{{\rm n}-1-10} & : n \neq 11 \\ (2({\rm n}-1))^{-1} & : n = 11 \end{cases}$$

The limit on the contamination becomes more stringent when more redshift bins are included in the analysis. Figure 2.1 gives the upper limit on f as a function of the number of redshift bins. The contamination must be less than or equal to these values so that the maximum error made in determining $\omega_{\rm ii}^{\rm T}$ via Equation (2.15) is 10 per cent.

With these limitations in mind we adopt the pairwise analysis throughout the rest of this work. A standard maximum likelihood procedure is used to estimate the contamination fractions by measuring the angular correlation functions at multiple scales and fitting them with Equation (2.14). Measurements of the angular correlation functions at different scales are not independent, thus the errors must be described by a covariance matrix. We use a bootstrapping technique to construct the covariance matrix, and include an additional source of error which we refer to as the clustering covariance matrix (Van Waerbeke, 2010). We refer the reader to Section 2.6 for the details of the covariance matrix and the likelihood method.

2.4 Application to a simulated galaxy survey

The Millennium Simulation tracks the hierarchical growth of dark matter structure from a redshift of 127 to the present. The simulation volume is a periodic box of 500 Mpc $\rm h^{-1}$ on a side, containing 2160^3 particles each with a mass of $8.6 \times 10^8 {\rm M}_\odot$. The simulation assumes a concordance model ΛCDM cosmology, $\Omega_{\rm m} = \Omega_{\rm dm} + \Omega_{\rm b} = 0.25$, $\Omega_{\rm b} = 0.045$, $\Omega_\Lambda = 0.75$, $h = 0.73$, $n = 1$ and $\sigma_8 = 0.9$, though the details of the cosmology do not affect any of the results presented here. A complete description of the Simulation is presented in Springel et al. (2005) and references therein.

We employ mock observational catalogues constructed from the Millennium Simulation to test

the method described in Section 2.3. The catalogues consist of six square pencil beam fields of 1.4 degrees on a side, containing a total of 28.7 million galaxies at redshifts less than 4.0. A full description of the catalogues is given by Kitzbichler & White (2007). To identify the mock catalogues, Kitzbichler & White (2007) give each a label 'a' through 'f'. We maintain this notation throughout the current work when distinguishing between the catalogues.

It is important to note that these mock observational catalogues do not include the effects of lensing, and in particular they do not account for weak lensing magnification. Since magnification will create an angular correlation between high and low redshift, it is relevant to the current work and will be addressed in Section 2.8.

We first present the results when applying the two-bin analysis to the mock observational catalogues, and discuss the effects of bin width and galaxy density. We then apply the global pairwise analysis demonstrating its ability to recover the true number of galaxies in a redshift bin, and the true average redshift of galaxies in each redshift bin.

2.4.1 Null test

Since there is no contamination between redshift bins in the simulated data, the angular cross-correlation function between any two redshift bins ought to be consistent with zero. To test this, two redshift bins are chosen so that there is, approximately, an equal number of galaxies in each; the bins are $0.3 < z_1 < 0.5$ which contains 69,139 galaxies per field on average and $0.8 < z_2 < 0.85$ which contains 50,575 galaxies per field on average.

The top left panel of Figure 2.2 shows the result of measuring the angular correlation functions on each of the six fields. The auto-correlation functions for a given redshift slice have different amplitudes due to the presence of different structures in each of the fields. The insert shows a zoom of the cross-correlation functions, and the errors include a bootstrapping term as well as the clustering term (see Section 2.6). The clustering term dominates the error at all but the smallest scales, providing a relatively uniform source of noise which behaves like a constant shift in the cross-correlation function. As seen in the top left panel of Figure 2.2; the six fields provided by the mock catalogues are slightly biased towards a positive shift. However, the clustering term is symmetric about zero and with enough measurements should have no bias. To test this we measured the cross-correlation between all combinations of high- and low-redshift bins, so that the high-redshift bin from field 'a' is cross-correlated with the low redshift bin of fields 'a', 'b', 'c', 'd', 'e' and 'f'. The resulting 36 cross-correlations are displayed in Figure 2.3, where the solid (red) line shows the average value which is consistent with zero at all scales.

The contamination fractions are estimated from the measured angular correlation functions as described in Section 2.6. The resulting likelihood is presented in the bottom left panel of Figure 2.2. Zero contamination is consistent with the results, and contamination in excess of $\sim$2 per cent can be ruled out at the 99.9 per cent confidence level.

We conclude that, as expected, there is no significant angular cross-correlation between widely separated redshift bins in the Millennium Simulation. Furthermore our method of estimating the contamination performs well in this case, indicating that there is less than $\sim$1 per cent contamination at the

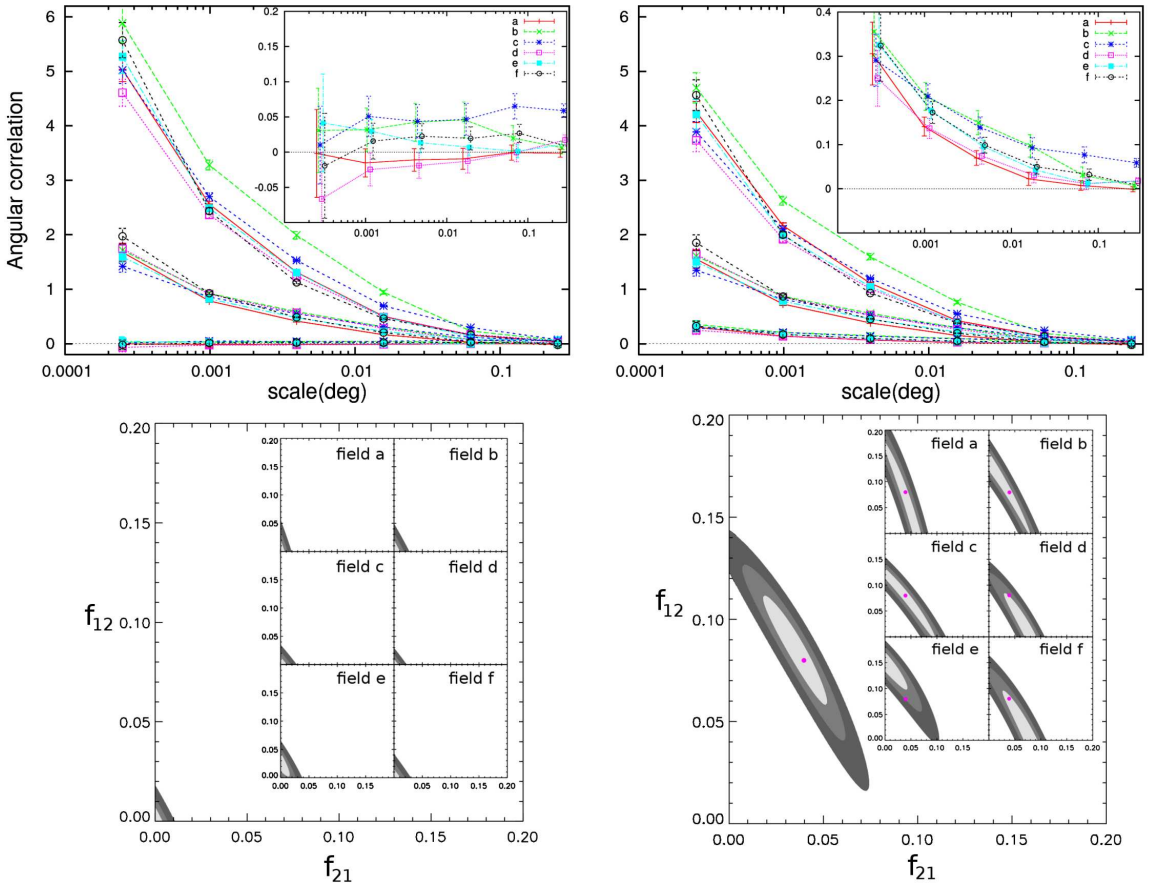

Figure 2.2: The left side panels are for the case of no contamination, the right side panels are for a contamination of $f_{12} = 0.08$ and $f_{21} = 0.04$. **Top row:** The angular auto- and cross-correlation functions for each test contamination of the mock catalogues. Each of the six fields is plotted with a different line style (colour). The higher redshift bin, $0.8 < z_2 < 0.85$, has a larger amplitude than the lower redshift bin, $0.3 < z_1 < 0.5$. A zoom of the cross-correlation is given in the insert, and the legend therein identifies which line style applies to each of the six mock catalogues. The error bars on the auto-correlation functions come from bootstrapping the catalogue. The cross-correlation functions also include the contribution from the clustering covariance (refer to Section 2.6 for details). The cross-correlation is consistent with zero for the case of zero contamination (left panel) and deviates significantly from zero for the contaminated case (right panel) **Bottom row:** The likelihood region of the contamination fractions. The shaded regions denote the 68, 95 and 99.9 per cent confidence areas, with increasing darkness indicating increasing significance. The contours result from summing the log-Likelihoods for the six individual fields. The insert shows the result for each of the individual fields. The input contamination in the right panel is marked with a dot.

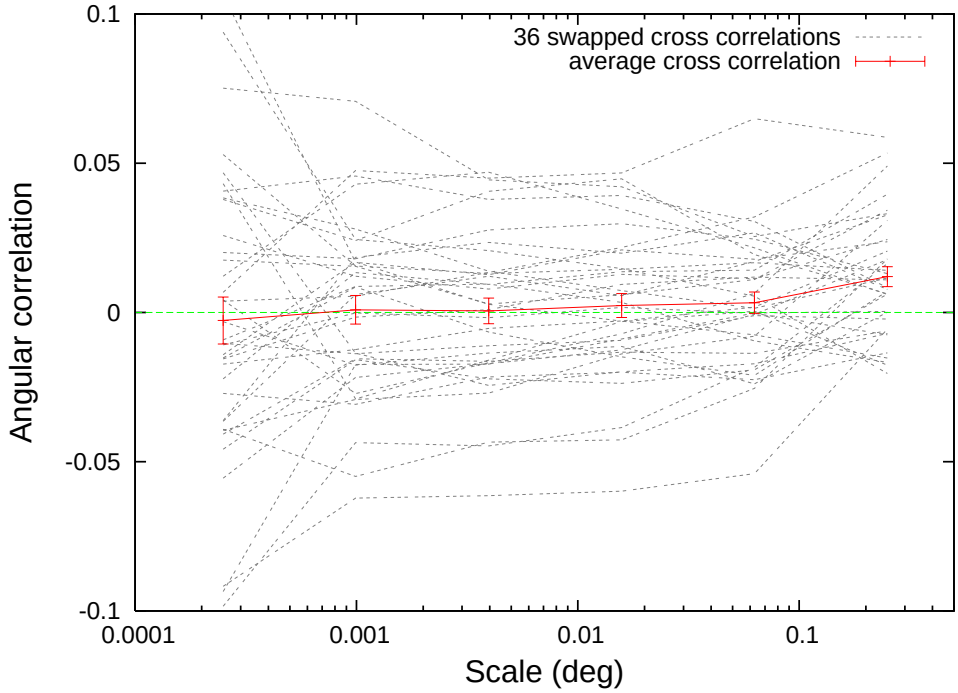

Figure 2.3: The 36 dotted (gray) lines are all of the possible cross-correlations found by swapping the high and low redshift bins between fields. The dashed (green) line shows y=0. The solid (red) line is the average of the 36 cross-correlations, and the error bar is the standard error. The average is consistent with zero, showing that the clustering term does not bias the cross-correlation.

68 per cent confidence level.

2.4.2 Artificial contamination

Contamination was added to each field by randomly shifting galaxies between the redshift bins. The contamination fractions were taken to be $f_{12} = 0.08$ and $f_{21} = 0.04$. The measured angular correlations are shown in the top right panel of Figure 2.2 where a cross-correlation signal is clearly seen. The recovered likelihood for the contamination fractions is shown in the bottom right panel of Figure 2.2. The input value lies in the 68 per cent confidence region, and zero contamination ($f_{12} = f_{21} = 0.00$) is ruled out well beyond the 99.9 per cent confidence level.

The parameter space is highly degenerate; contamination in either direction (low to high redshift, or vice versa) increases the cross-correlation amplitude. Degeneracy breaking comes from the ability to probe the shape of the auto-correlation functions. If only one scale is used, the parameters are completely degenerate. Likewise if the auto-correlation of each redshift bin has the same shape (for example, if they are power laws with the same slope), then the parameters are completely degenerate.

Each field has independent structure, causing variations in the measured correlation functions. It is important that this is not considered a source of sample variance. Any features in the auto-correlation

function will also be present in the cross-correlation function if contamination is present. The inserts in the bottom panels of Figure 2.2 show the parameter constraints obtained by analysing each field independently. These contours are combined, yielding a better measurement of the contamination.

We have repeated this procedure for several different contamination fractions (f_{12}, f_{21}) including (0.0,0.02), (0.15,0.0), (0.0,0.15) and (0.0,0.45). In all cases the input contamination was recovered within the 68 per cent confidence region.

2.4.3 Effect of galaxy density and redshift bin width

Each of the six mock catalogues of Kitzbichler & White (2007) contain on average 4.79×10^6 galaxies out to redshift 4. The two redshift bins used thus far, $0.3 < z < 0.5$ and $0.80 < z < 0.85$, were chosen to be well separated in redshift and contain roughly equal numbers of galaxies. On average there are $\sim 6 \times 10^4$ galaxies per redshift bin per field, corresponding to a density of 8.5 arcmin^{-2}.

In order to test the effects of object density, galaxies were randomly removed from the mock catalogues which were then contaminated with $f_{12} = 0.08$ and $f_{21} = 0.04$. Densities of 4.3 and 1.7 arcmin^{-2} were tested. We found that lower densities yielded only marginally weaker constraints compared to those in the bottom right panel of Figure 2.2.

Reducing object density has a greater effect on constraints from individual fields; much of the constraining power results from the combination of the six fields. As long as there are sufficient galaxies to measure the angular correlation function accurately, the contamination can be constrained. There is a disproportionately small change to the contours when the density is reduced since we are not dominated by Poisson noise.

If more narrow redshift bins are taken, for example $0.3 < z < 0.4$ and $0.8 < z < 0.82$, then the constraints on the contamination improve considerably. Each bin contains 24,451 galaxies per field on average with a corresponding density of 3.5 arcmin^{-2}. Figure 2.4 shows the parameter constraints when a contamination of $f_{12} = 0.08$ and $f_{21} = 0.04$ is used, the constraints for wider bins are over-plotted for comparison. The narrow bins offer much tighter constraints despite the density being lowered as a result of the smaller bin size.

The angular correlation function for narrower redshift bins is more distinct since a larger fraction of galaxies are physically clustered with each other. Making the bins wider dilutes the sample with galaxies that are not clustered, reducing the correlation. Arbitrarily narrow bins will cluster strongly, providing a good measure of the correlation function, but will suffer due to low number densities. More work is needed to determine an optimal trade-off between these parameters.

2.4.4 Global pairwise analysis

We divide the data into the following redshift bins: [0.0,0.5], (0.5,0.8], (0.8,1.1] and (1.1,1.5], which we label z_1, z_2, z_3 and z_4 respectively. The average number of galaxies in each bin from low to high redshift is: 60804, 94907, 63707, 55290. In order to test the global pairwise analysis, contamination fractions between all redshift bins must be specified. In lieu of a completely random contamination matrix we have taken one which is similar to the contamination we measure for the CFHTLS-Deep

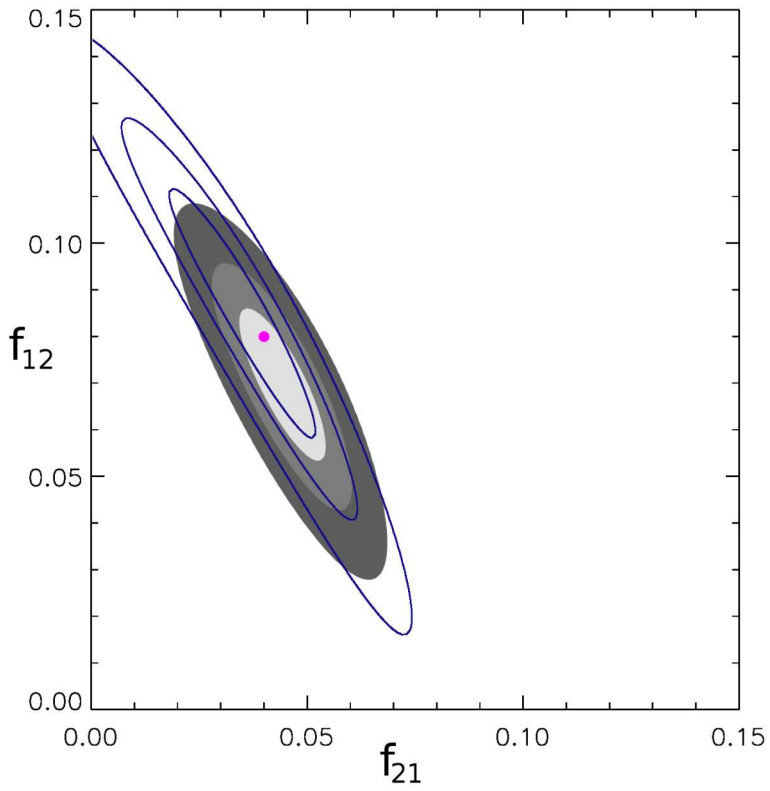

Figure 2.4: Constraints on the contamination fractions for an input model of $f_{12} = 0.08$ and $f_{21} = 0.04$, which is marked with a dot. The lined contours show the results for the wider redshift bins; $0.3 < z < 0.5$ and $0.80 < z < 0.85$. The filled contours give the constraints for more narrow redshift bins; $0.3 < z < 0.4$ and $0.8 < z < 0.82$. Narrow redshift bins provide tighter constraints despite the decreased number density.

fields in Section 2.5. Since the measured contamination in the Deep fields appears to be on the extreme end of the global pairwise approximation we have reduced some of the contamination values, though the adopted contamination matrix remains a strong test of the global pairwise approximation. The contamination matrix used is,

$$f_{\rm ij} = \begin{pmatrix} 0.70 & 0.01 & 0.10 & 0.10 \\ 0.10 & 0.925 & 0.15 & 0.04 \\ 0.05 & 0.06 & 0.70 & 0.08 \\ 0.15 & 0.005 & 0.05 & 0.78 \end{pmatrix}, \tag{2.19}$$

where $f_{\rm ij}$ refers to the entry in the i$^{\rm th}$ column and the j$^{\rm th}$ row. Galaxies are chosen randomly to be moved between redshift bins such that Equation (2.6) is satisfied.

The result of measuring the angular auto-correlation function for each redshift bin and the cross-correlation function for each pair of redshift bins, as measured from field 'a', is presented in the left panel of Figure 2.5. One hundred bootstraps of the contaminated catalogues were constructed, from

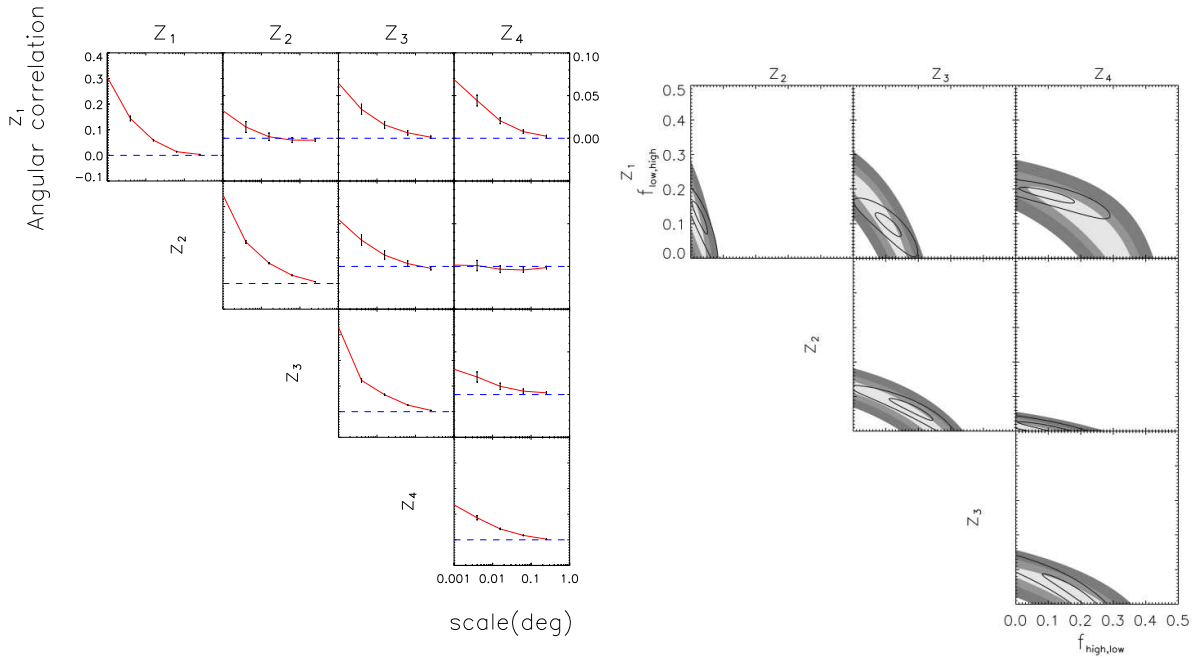

Figure 2.5: The redshift bins are: z_1 =[0.0,0.5], z_2 =(0.5,0.8], z_3 =(0.8,1.1] and z_4 =(1.1,1.5]. **Left panel:** The angular correlation functions as measured for field 'a'. The subplots along the diagonal contain the auto-correlation for each redshift bin. The off-diagonal subplots contain the cross-correlation function between two bins. The redshift bin labels at the top of each column and to the left of each row denote which bins are involved in the given correlation function. The vertical scale on the top-left subplot applies to all subplots on the diagonal, the scale on the top-right subplot applies to all off-diagonal subplots. **Right panel:** The filled contours depict the parameter constraints estimated for field 'a'. The shading from light to dark represents the 68, 95 and 99.9 per cent confidence levels. The lined contours are the result of combining the parameter constraints from each individual field, for clarity we only show the 68 and 99.9 per cent confidence levels. Each subplot contains the constraints for a pair of contamination fractions. The x-axis is taken to be the contamination fraction from the high redshift bin to the low redshift bin and the y-axis is the reverse. The two bins contributing to each subplot are indicated by redshift bin labels at the top of each column and to the left of each row.

which the angular correlation functions were measured. These hundred measurements were used to estimate the covariance matrix, applying the correction described in Hartlap et al. (2007). For the auto-correlation functions (subplots on the diagonal) this is the only source of error, the cross-correlation functions (off-diagonal subplots) have the clustering term as an additional source of error.

The pairwise analysis is applied to each pair of redshift bins yielding constraints on the contamination between each pair. The constraints on the contamination fractions for field 'a' are given by the filled contours in the right panel of Figure 2.5. Each of the six fields is considered independent, and their constraints on the contamination fractions are combined by multiplying the likelihoods together. These combined constraints are given as lined contours, over-plotted in Figure 2.5.

The pairwise analysis does not impose global constraints since each pair of bins is considered in-

dependently. Not all combinations of contamination fractions suggested by Figure 2.5 will produce a consistent picture. For example, it is possible to select points from each plot in the right most column such that more than 100 per cent of galaxies from the highest redshift bin are contaminating other bins. To get a sense of which solutions are globally consistent, and what a typical global solution looks like, we employ a Monte-Carlo method. We first randomly sample 1000 points from each of the pairwise likelihood regions such that the density of points reflects the underlying probability distribution. Taking a random point from each pairwise analysis will uniquely specify a realization of the global contamination, i.e. the matrix of Equation (2.6). Thus a realization of the true redshift distribution can be determined. We then impose two constraints: the true number of galaxies in a bin cannot be less than zero, and no more than 100 per cent of galaxies can be scattered from a single bin. This procedure is repeated until 200,000 admissible realizations are found.

We have verified that the admissible realizations of the contamination fractions are representative of the full probability distributions from which they are drawn. It is not the case that the global constraints exclude particular regions of parameter space. This is checked by constructing a probability distribution for each contamination fraction from the 200,000 measurements and comparing it to the marginalised PDF obtained from the likelihood contours. In all cases the two agree with each other. For the combined case, the estimated contamination fractions (taken to be the average value) with 68 per cent confidence levels are as follows,

$$f_{\rm ij} = \begin{pmatrix} 0.60^{+0.04}_{-0.05} & 0.03^{+0.02}_{-0.02} & 0.11^{+0.03}_{-0.03} & 0.11^{+0.05}_{-0.06} \\ 0.12^{+0.04}_{-0.03} & 0.89^{+0.04}_{-0.03} & 0.18^{+0.05}_{-0.04} & 0.09^{+0.05}_{-0.05} \\ 0.10^{+0.02}_{-0.02} & 0.07^{+0.02}_{-0.02} & 0.66^{+0.06}_{-0.06} & 0.16^{+0.07}_{-0.03} \\ 0.18^{+0.02}_{-0.01} & 0.01^{+0.00}_{-0.01} & 0.05^{+0.02}_{-0.05} & 0.65^{+0.09}_{-0.09} \end{pmatrix}. \tag{2.20}$$

This result agrees extremely well with the input contamination matrix of Equation (2.19), with 11 of the 16 contamination fractions, 69 per cent, agreeing within the 68 per cent error estimate. As expected the pairwise analysis slightly overestimates the level of contamination, which is probably more pronounced in this case since the contamination matrix is so aggressive. The contamination fractions on the off-diagonal are overestimated since the pairwise method assumes that the observed angular cross-correlation function is only due to contamination between the two considered redshift bins. However, the pair of redshift bins considered will each be contaminated from other redshift bins, leading to a cross-correlation between the redshift bins which is not dependent on the scattering of redshifts between them. Note that the diagonal elements in Equation (2.20) are smaller than for the input contamination matrix, which is also a result of overestimating the contamination fractions (see Equation 2.4).

Using Equation (2.6) allows us to estimate the true number of galaxies in each redshift bin for each of the 200,000 realizations of the contamination matrix. This is done for each of the six fields as well as the combined case, the probability distributions for the estimated true number of galaxies are presented in Figure 2.6. The cross hashed regions denote the 68 per cent confidence level. Since this is simulated data with artificial contamination we know what the actual uncontaminated number of galaxies is for a given bin, this is over-plotted as a dashed vertical line, for clarity we will refer to this as the true number

of galaxies, and to our attempt to recover this as the estimated true number of galaxies. The solid vertical line shows the observed number of galaxies; i.e., the number of galaxies after contamination. For the combined case the number of observed galaxies is taken to be the average number of galaxies from the 6 fields. The global pairwise analysis does a good job of recovering the true number of galaxies in each bin, although in many cases the observed number is also an acceptable fit, owing to the small contamination fractions and the fact that the redshift distribution was nearly flat to begin with. Focusing on the combined result for z_2 we see that the true number of galaxies is a significantly better fit to the estimated true number than the observed number.

It is also possible to estimate the true average redshift of a redshift bin. The true average redshift is defined as the average redshift of galaxies in a redshift bin which has contamination. With real world data it is the same as asking what the average spectroscopic redshift is in a given photometric redshift bin. This is straightforward to calculate for a simulated survey since we know for each redshift bin the fraction of galaxies that came from each other redshift bin. The true average redshift is given by:

$$\bar{z}_{\rm i}^{\rm T} = \frac{1}{N_{\rm i}^{\rm o}} \sum_{\rm k=1}^{\rm n} \bar{z}_{\rm k}^{\rm uncontam} f_{\rm ki} N_{\rm k}^{\rm T}, \tag{2.21}$$

where $\bar{z}_{\rm k}^{\rm uncontam}$ is the uncontaminated average redshift of a galaxy in bin k, and $f_{\rm ki}N_{\rm k}^{\rm T}$ is the number of galaxies in bin i from bin k. In general the uncontaminated average redshift of a bin is not known since it requires knowledge of the true redshift of each galaxy. If we assume that the contamination does not significantly alter the shape of the redshift distribution at the sub-redshift bin level, then the average redshift of a bin will not be changed by contamination. Therefore we estimate the average redshift of an uncontaminated redshift bin ($\bar{z}_{\rm k}^{\rm uncontam}$) by the average redshift of the contaminated redshift bin.

For a given field, each of the 200,000 realizations of the contamination matrix yields an estimate of the true average redshift for each redshift bin. Thus a probability distribution function is constructed for each bin and each field, including the combined case, this result is presented in Figure 2.7. The solid vertical lines show the average redshift with no contamination, this is what one would measure to be the average redshift if the effects of contamination were ignored. The dashed vertical lines show the true average redshift for each bin, which we can measure directly here because we're working with simulated data with known contamination. Our method does a very good job of recovering the true average redshift. The lowest and highest redshift bins suffer the most, this is expected since they can only be contaminated by galaxies that are either higher or lower in redshift respectively, whereas the middle bins are contaminated by galaxies which are both higher and lower than the average, allowing for a cancellation of the effect.

We have demonstrated that the global pairwise analysis can be used to reliably recover small contaminations between redshift bins. Using the Monte Carlo approach detailed in this section we have shown that we are able to estimate the true redshift distribution, and the true average redshift within a bin.

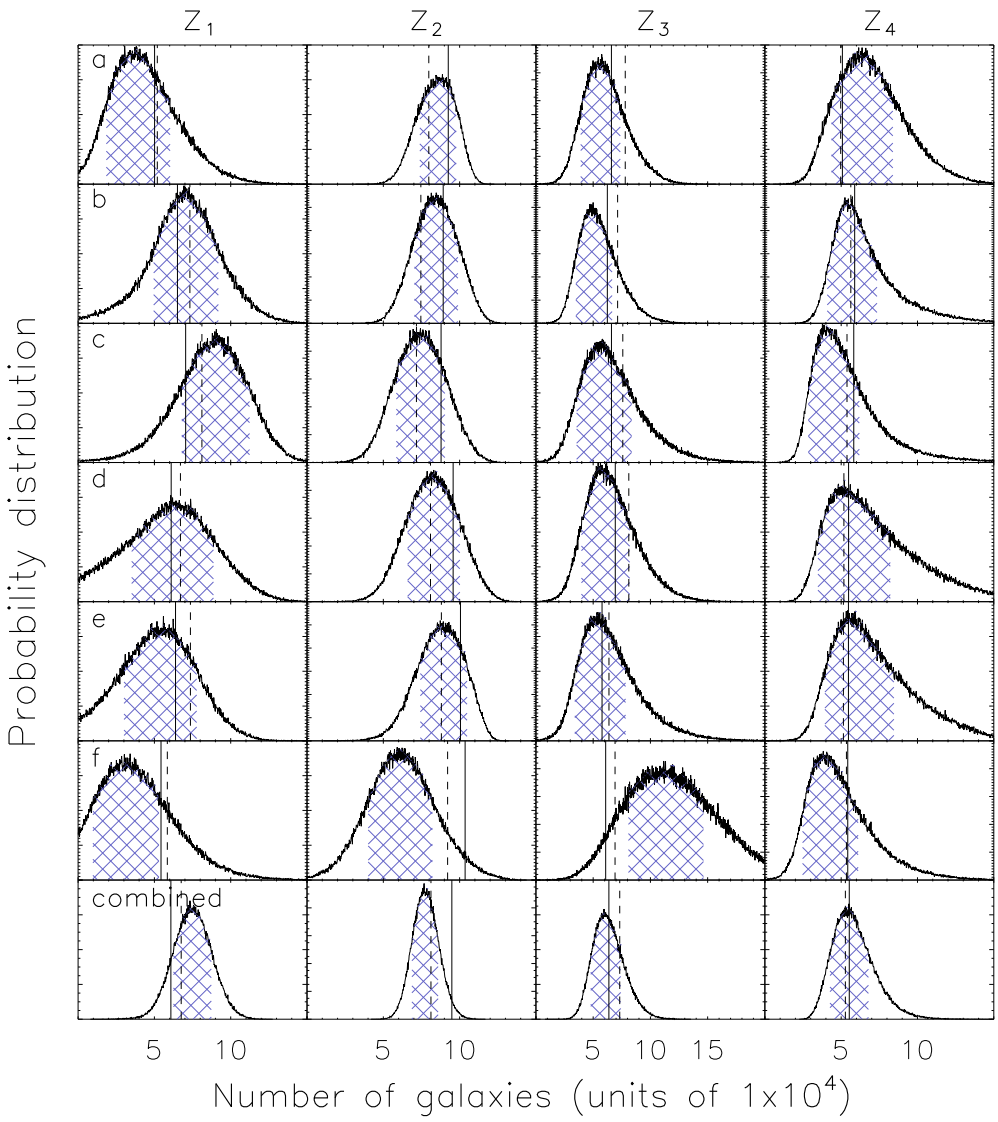

Figure 2.6: Millennium Simulation: This demonstrates the ability of the global pairwise analysis to estimate the true (uncontaminated) redshift distribution. The redshift bins are: z_1 =[0.0,0.5], z_2 =(0.5,0.8], z_3 =(0.8,1.1] and z_4 =(1.1,1.5]. The x-axis is the number of galaxies in units of 1×10^4. The y-axis is the probability, which has been scaled differently in each subplot for clarity. The histograms are the probability distribution of the true number of galaxies, and the cross-hashing denotes the 68 per cent confidence region. Each row of subplots is the result from one of the Millennium Simulation fields. The bottom row is the result when the constraints on the contamination fractions for each field are combined (see Figure 2.5). Please note that the bottom row is not a direct combination of the results from the other rows. Each column represents a redshift bin, as labelled. The solid vertical line in each subplot indicates the observed number of galaxies. The dashed vertical line is the true number of galaxies.

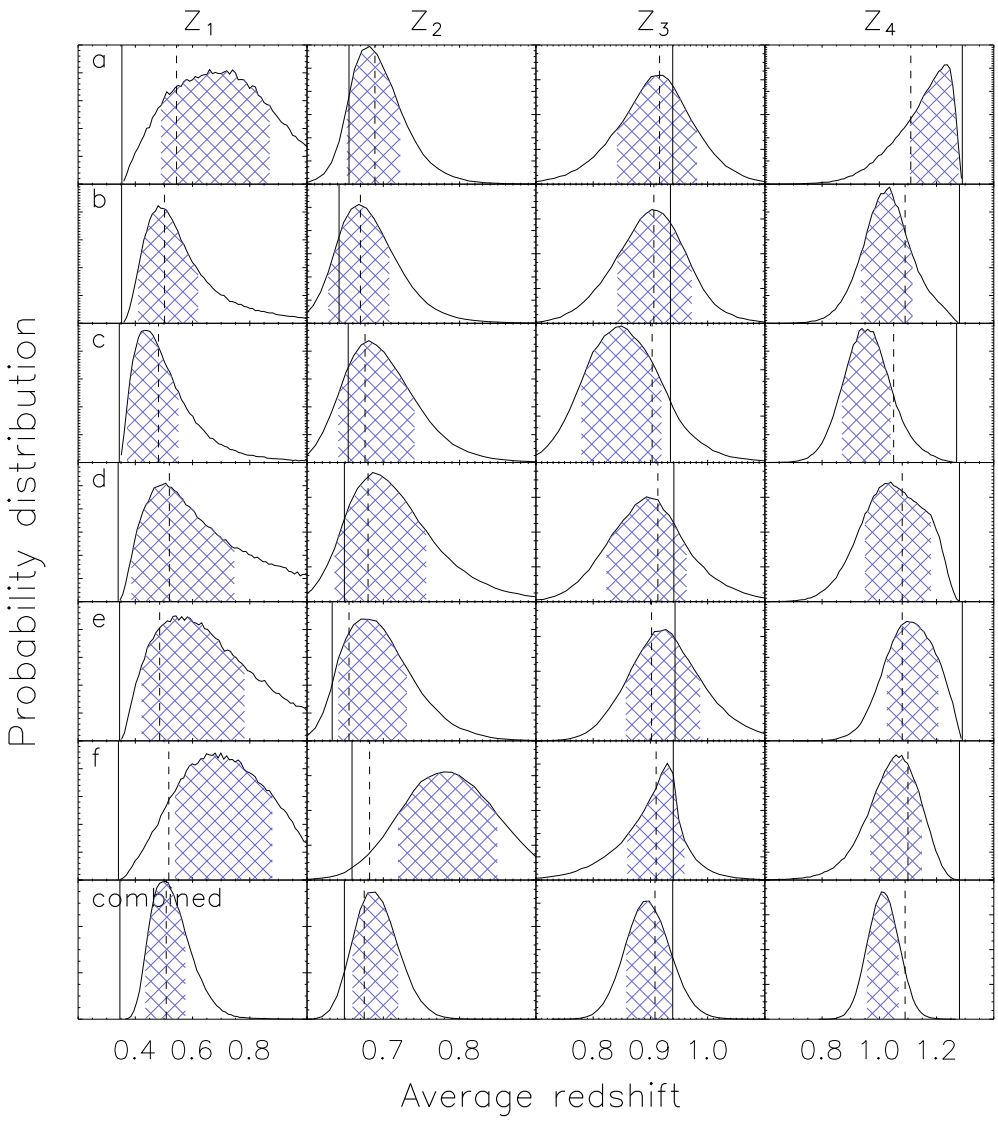

Figure 2.7: Millennium Simulation: The global pairwise analysis is used to estimate the true average redshift of each photometric redshift bin. The redshift bins are: z_1 =[0.0,0.5], z_2 =(0.5,0.8], z_3 =(0.8,1.1] and z_4 =(1.1,1.5]. The x-axis is the average redshift. The y-axis is the probability, which has been scaled differently in each subplot for clarity. The histograms are the probability distribution of the average redshift, and the cross-hashing denotes the 68 per cent confidence region. Each row of subplots is the result from one of the Millennium Simulation fields. The bottom row is the result when the constraints on the contamination fractions for each field are combined (see Figure 2.5). Please note that the bottom row is not a direct combination of the results from the other rows. Each column represents a redshift bin, as labelled. The solid vertical line in each subplot indicates the average redshift as measured from the uncontaminated catalogue. The dashed vertical line is the true average redshift.

2.5 Application to a real galaxy survey

The Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) is a joint Canadian-French program designed to take advantage of Megaprime, the CFHT wide-field imager. This 36-CCD mosaic camera has a 1×1 degree field of view and a pixel scale of 0.187 arcseconds per pixel. The deep component consists of four one-square-degree fields imaged with five broad-band filters: u^*, g', r', i', and z'. The fields are designated D1, D2, D3 and D4, and are centred on RA;DEC coordinates of 02 26 00; -04 30 00, 10 00 28; +02 12 21, 14 19 28; +52 40 41 and 22 15 32; +17 44 06 respectively

We use the deep photometric redshift catalogues from Ilbert et al. (2006) who have estimated redshifts for the T0003 CFHTLS-Deep release. The full photometric catalogue contains 522286 objects, covering an effective area of 3.2 deg^2. A set of 3241 spectroscopic redshifts with $0 \leq z \leq 5$ from the VIRMOS VLT Deep Survey (VVDS) were used as a calibration and training set for the photometric redshifts. The resulting photometric redshifts have an accuracy of $\sigma_{(z_{phot}-z_{spec})/(1+z_{\rm spec})} = 0.043$ for $i'_{\rm AB}$ = 22.5 - 24, with a fraction of catastrophic errors of 5.4 per cent.

In this work we consider galaxies with $21.0 < i' < 25.0$ and divide the data into the following redshift bins: [0.0,0.2], (0.2,1.5], (1.5,2.5] and (2.5,6.0], which we label z_1, z_2, z_3 and z_4 respectively. The average number of galaxies in each bin from low to high redshift is: 5772, 96019, 15546, 5315. These redshift bins are chosen to isolate the low confidence redshifts and to probe the bump in the redshift distribution near redshift 3 (see Figure 2.8). We would like to know if this feature is an artefact of the photometric redshifts or if it corresponds to a physical over-density of galaxies at this redshift.

Adopting the definition of catastrophic error used by Ilbert et al. (2006), $\Delta z > 0.15(1+z)$, we measure the fraction of galaxies with catastrophic errors in each bin. From low to high redshift we find: 23.2, 12.4, 33.2 and 23.1 per cent. It is difficult to interpret these values in the context of the contamination fractions. Photometric redshifts with large errors do not necessarily contaminate other redshift bins. Furthermore Δz does not take into account degeneracies in the spectral template fitting that allow for alternative solutions to the galaxy's redshift. These catastrophic errors give some indication of the leakage between adjacent redshift bins but cannot account for those galaxies which are completely misclassified.

2.5.1 Applying the global pairwise analysis

We apply the two-bin analysis between each pair of redshift bins, and for each of the four Deep fields. The measured angular correlation functions for D1 are presented in Figure 2.9. The covariance matrices are estimated by bootstrapping the catalogues 100 times, and applying the correction described by Hartlap et al. (2007). For the cross-correlation covariance matrix we also calculate the clustering covariance described in Section 2.6. Parameter constraints are estimated for each cross-correlation, those for D1 are presented in Figure 2.9. The other three fields have very similar angular correlation functions and parameter constraints. The parameter constraints for all of the fields can be combined by treating them as statistically independent and multiplying their likelihoods together, yielding tighter constraints (lined contours in Figure 2.9).

We constructed 200,000 realizations of a globally consistent contamination matrix, as detailed in

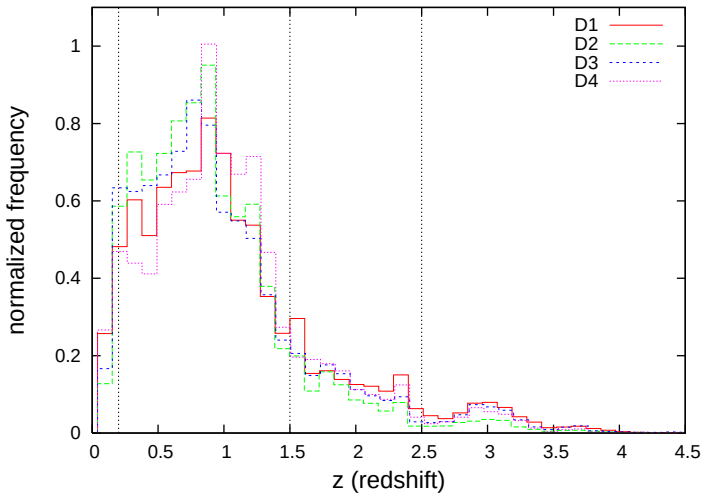

Figure 2.8: Finely binned redshift distribution for each of the four deep fields. The vertical lines denote the binning adopted in our global pairwise analysis: The redshift bins are: z_1 =[0.0,0.2], z_2 =(0.2,1.5], z_3 =(1.5,2.5] and z_4 =(2.5,6.0]. Note the bump near redshift 3.

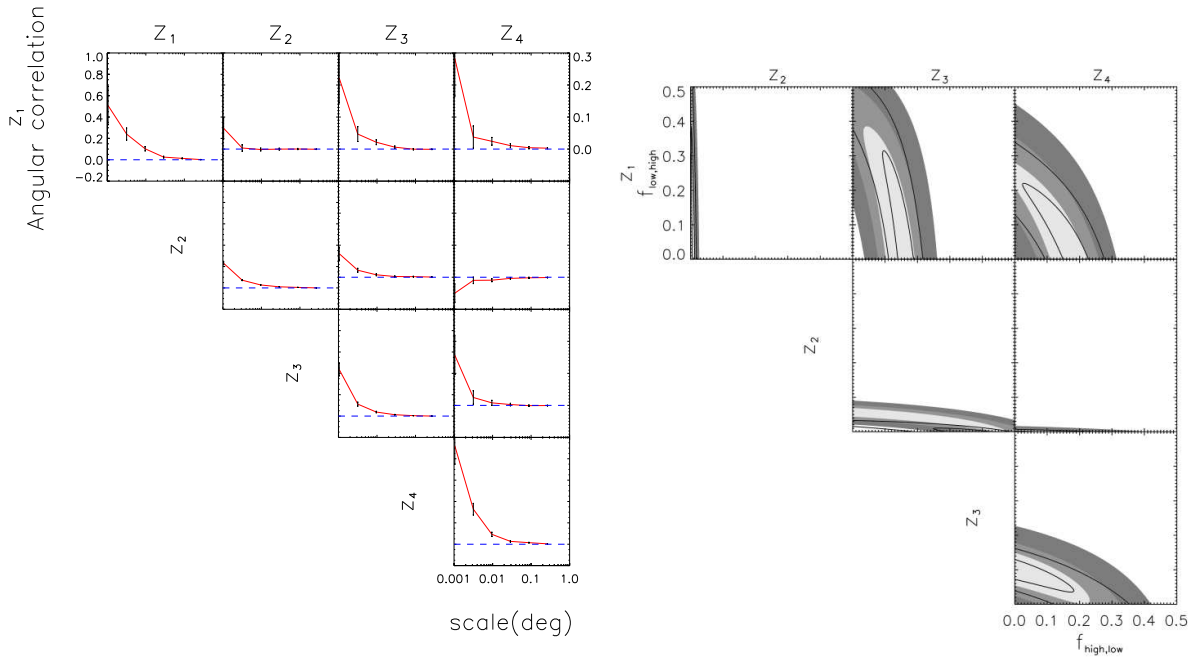

Figure 2.9: Same as caption to Figure 2.5 but with field D1 in place of field 'a'. The redshift bins are: z_1 =[0.0,0.2], z_2 =(0.2,1.5], z_3 =(1.5,2.5] and z_4 =(2.5,6.0].

Section 2.4.4, and verified that the admissible realizations are representative of the full probability distributions they are drawn from. The contamination matrix estimated from the combined constraints is

$$f_{ij} = \begin{pmatrix} 0.50^{+0.17}_{-0.17} & 0.006^{+0.004}_{-0.003} & 0.13^{+0.04}_{-0.02} & 0.13^{+0.07}_{-0.04} \\ 0.18^{+0.06}_{-0.18} & 0.979^{+0.014}_{-0.006} & 0.31^{+0.08}_{-0.04} & 0.04^{+0.01}_{-0.04} \\ 0.16^{+0.05}_{-0.16} & 0.006^{+0.001}_{-0.006} & 0.53^{+0.05}_{-0.10} & 0.09^{+0.03}_{-0.09} \\ 0.11^{+0.03}_{-0.11} & 0.001^{+0.000}_{-0.001} & 0.07^{+0.04}_{-0.02} & 0.74^{+0.10}_{-0.10} \end{pmatrix}, \quad (2.22)$$

where the entries represent the average value calculated from their probability distribution. The maximum contamination fraction for four redshift bins presented in Figure 2.1 is $\sim$0.12 which is smaller than some of the contamination fractions found here (or seen in Figure 2.9). It is possible that this is an indication that the pairwise analysis does not hold for these data; however, several assumptions were made in deriving the maximum contamination fraction which do not hold here. We assumed that all contamination fractions have the same value; this is not the case here, and although some tend to be larger than 0.12, some are very close to zero, and several have large errors encompassing zero. We also assumed that there were equal numbers of galaxies in each bin. The ratio of the number of galaxies between the pair of bins enters into the expression for the observed cross-correlation. Since the first and last redshift bins have far fewer galaxies, a large contamination fraction from one of these bins represents only a small number of galaxies. We have also demonstrated our ability to recover the contamination fractions from a similarly aggressive contamination matrix in Section 2.4.4. For these reasons we believe that the global pairwise analysis remains a good approximation here.

The probability distribution of the true number of galaxies for each redshift bin and each field is presented in Figure 2.10; the cross-hatched regions indicate 68 per cent confidence. The observed number of galaxies in each bin is denoted by a vertical line. The bottom row contains the result when the constraints on the contamination fractions for each of the four fields are combined. The smallest fractional change is for z_2 which is the high confidence photometric redshift bin as defined by Ilbert et al. (2006). The peak of the probability distribution indicates about a factor of two fewer galaxies in the highest redshift bin than are observed, suggesting that the bump seen in the photometric redshift distribution is an artefact of contamination. However, with only four square degrees of data, we are unable to rule out the existence of this feature.

The set of globally consistent realizations of the contamination can also be used to estimate the average redshift for each photometric redshift bin. We use Equation (2.21), and estimate the uncontaminated average redshift of each bin ($\bar{z}_k^{\rm uncontam}$) by the average of the photometric redshifts. Which is a good approximation as long as the shape of the observed redshift distribution within each bin is similar to that of the true redshift distribution.

The results are presented in Figure 2.11 which shows the probability distribution of the average redshift for each redshift bin and each field. Vertical lines show the average redshift for each bin measured from the photometric catalogue. It is clear that the smallest, and largest, redshift bins (z_1, and z_4) contain galaxies whose true average redshifts deviate significantly from the average redshift expected for those redshift bins. This suggests that many galaxies in bin z_1 are in fact from much higher redshifts.

Similarly galaxies in bin z_4 have a lower than expected average redshift.

2.6 Covariance and likelihood

Here we present the details of the maximum-likelihood method, and the covariance matrix used. Throughout this chapter we fit the observed angular cross-correlation between two redshift bins with the model described by Eq.(2.14). Since there are observational errors associated with the angular auto- and cross-correlation functions, we have grouped these quantities on the left hand side of Eq.(2.14), yielding:

$$\omega^{\rm o}_{\rm ij}(\theta)(f_{\rm ii}f_{\rm jj}+f_{\rm ij}f_{\rm ji})-\omega^{\rm o}_{\rm ii}(\theta)\frac{N^{\rm o}_{\rm i}}{N^{\rm o}_{\rm j}}f_{\rm ij}f_{\rm jj}-\omega^{\rm o}_{\rm jj}(\theta)\frac{N^{\rm o}_{\rm j}}{N^{\rm o}_{\rm i}}f_{\rm ji}f_{\rm ii}=0, \tag{2.23}$$

where the angular correlation functions are written as a function of scale θ. For simplicity let $\mathbf{F}$ represent the left hand side of the equation. We therefore seek to calculate the likelihood,

$$\mathscr{L}=\frac{1}{\sqrt{(2\pi)^s|\mathbf{C}|}}\exp\left[-\frac{1}{2}(\mathbf{F}-\mathbf{m})\mathbf{C}^{-1}(\mathbf{F}-\mathbf{m})^T\right], \tag{2.24}$$

where s is the number of angular scale bins, $\mathbf{m}$ is the model which is zero for all scales and $\mathbf{C}$ is the $s\times s$ covariance matrix. The covariance matrix is

$$C_{kl}=\langle F_kF_l\rangle, \tag{2.25}$$

where k and l denote the scales at which the angular correlation functions are measured. Expanding the above yields

$$\begin{aligned}
C_{kl} &= \langle\omega^{\rm o}_{\rm ij}(\theta_k)\omega^{\rm o}_{\rm ij}(\theta_l)\rangle(f_{\rm ii}f_{\rm jj}+f_{\rm ij}f_{\rm ji})^2 \\
&+ \langle\omega^{\rm o}_{\rm ii}(\theta_k)\omega^{\rm o}_{\rm ii}(\theta_l)\rangle\left(\frac{N^{\rm o}_{\rm i}}{N^{\rm o}_{\rm j}}f_{\rm ij}f_{\rm jj}\right)^2 \\
&+ \langle\omega^{\rm o}_{\rm jj}(\theta_k)\omega^{\rm o}_{\rm jj}(\theta_l)\rangle\left(\frac{N^{\rm o}_{\rm j}}{N^{\rm o}_{\rm i}}f_{\rm ji}f_{\rm ii}\right)^2 \\
&+ \langle\omega^{\rm o}_{\rm ij}(\theta_k)\omega^{\rm o}_{\rm ii}(\theta_l)\rangle 2(f_{\rm ii}f_{\rm jj}+f_{\rm ij}f_{\rm ji})\frac{N^{\rm o}_{\rm i}}{N^{\rm o}_{\rm j}}f_{\rm ij}f_{\rm jj} \\
&+ \langle\omega^{\rm o}_{\rm ij}(\theta_k)\omega^{\rm o}_{\rm jj}(\theta_l)\rangle 2(f_{\rm ii}f_{\rm jj}+f_{\rm ij}f_{\rm ji})\frac{N^{\rm o}_{\rm j}}{N^{\rm o}_{\rm i}}f_{\rm ji}f_{\rm ii} \\
&+ \langle\omega^{\rm o}_{\rm ii}(\theta_k)\omega^{\rm o}_{\rm jj}(\theta_l)\rangle 2f_{\rm ij}f_{\rm jj}f_{\rm ji}f_{\rm ii}.
\end{aligned} \tag{2.26}$$

Ideally the covariance matrix can be estimated directly from the data, but this requires many fields. It is not possible to do this for either the Millennium Simulation or the CFHTLS-Deep data sets which we consider in this work. An alternative is to use a bootstrapping method, wherein the data catalogue is resampled multiple times, and each resampled catalogue is used to measure the angular correlation

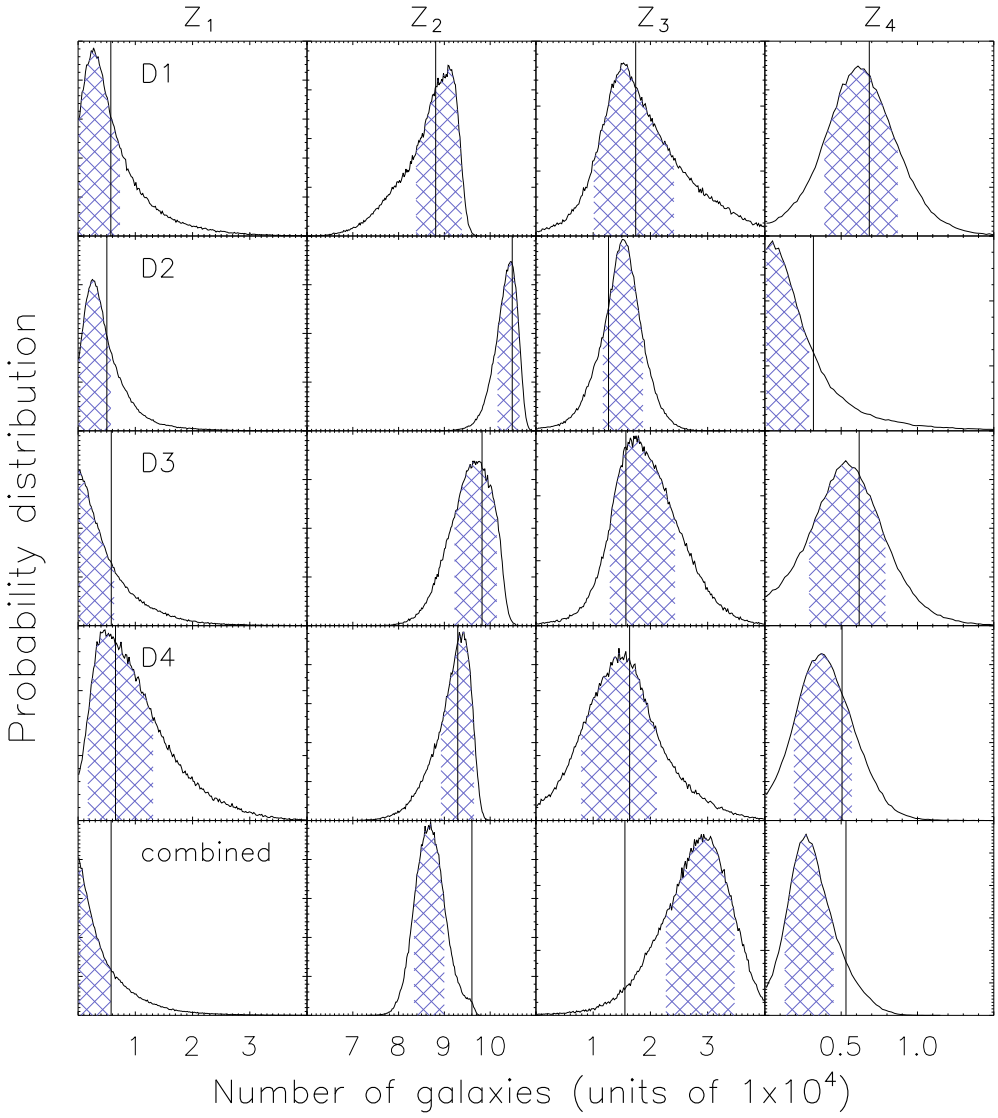

Figure 2.10: CFHTLS-Deep: This demonstrates the ability of the global pairwise analysis to reconstruct the true (uncontaminated) redshift distribution. The redshift bins are: z_1 =[0.0,0.2], z_2 =(0.2,1.5], z_3 =(1.5,2.5] and z_4 =(2.5,6.0]. The x-axis is the number of galaxies in units of 1×10^4. The y-axis is the probability, which has been scaled differently in each subplot for clarity. The histograms are the probability distribution of the true number of galaxies, and the cross-hashing denotes the 68 per cent confidence region. Each row of subplots is the result from one of the CFHTLS-Deep fields. The bottom row is the result when the constraints on the contamination fractions for each field are combined. Each column represents a redshift bin, as labelled. The vertical line in each subplot indicates the observed number of galaxies.

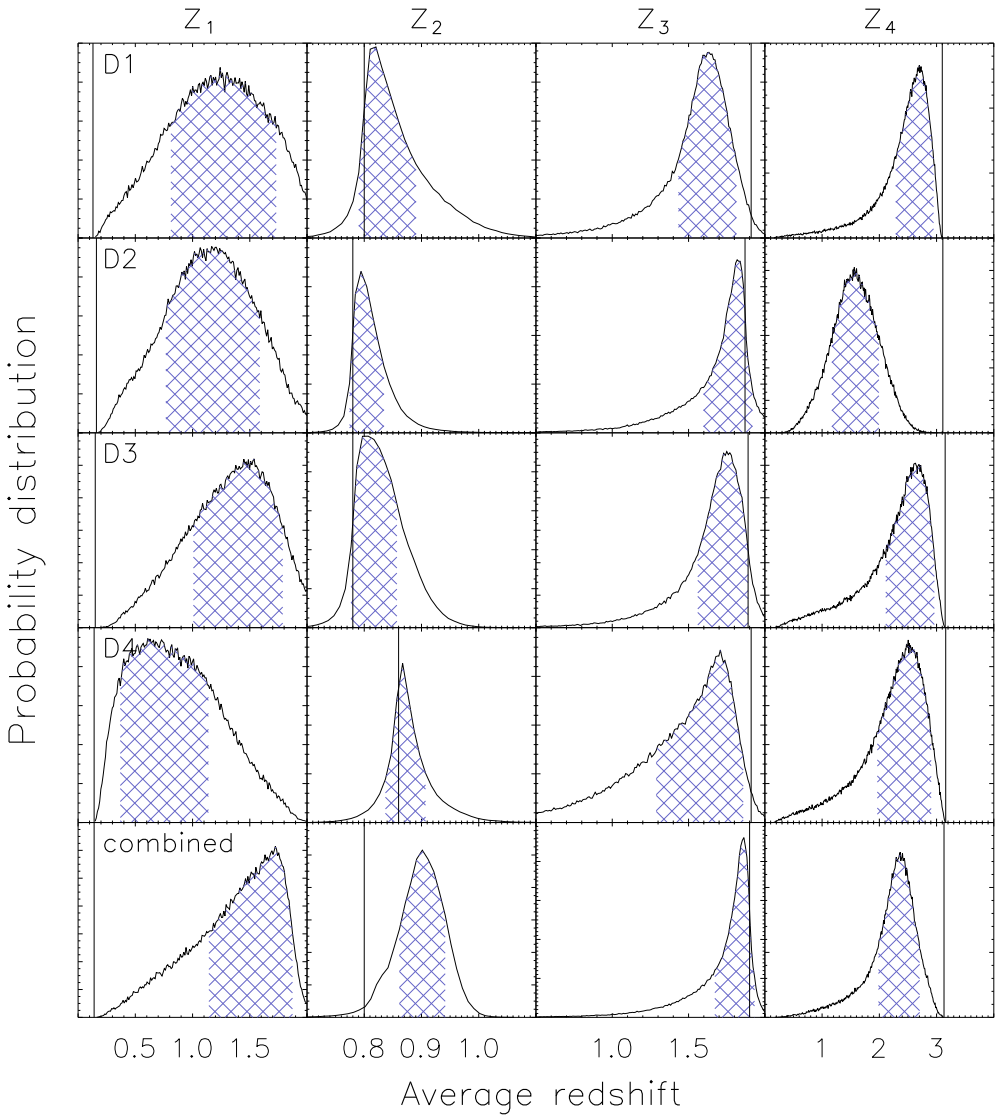

Figure 2.11: CFHTLS-Deep: The global pairwise analysis is used to estimate the true average redshift of each photometric redshift bin. The redshift bins are: z_1 =[0.0,0.2], z_2 =(0.2,1.5], z_3 =(1.5,2.5] and z_4 =(2.5,6.0]. The x-axis is the average redshift. The y-axis is the probability, which has been scaled differently in each subplot for clarity. The histograms are the probability distribution of the average redshift, and the cross-hashing denotes the 68 per cent confidence region. Each row of subplots is the result from one of the CFHTLS-Deep fields. The bottom row is the result when the constraints on the contamination fractions for each field are combined. Each column represents a redshift bin, as labelled. The vertical line in each subplot indicates the average redshift as measured from the photometric redshift catalogue.

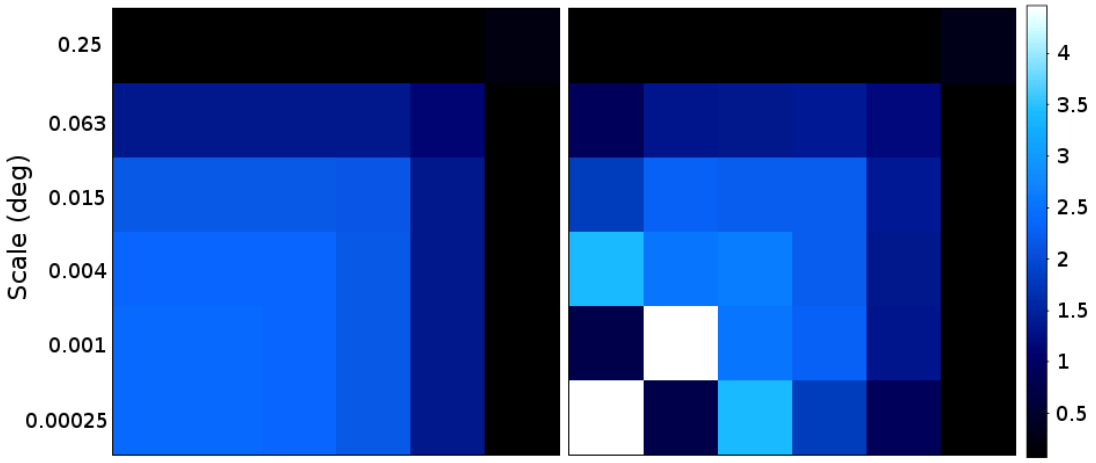

Figure 2.12: The covariance matrix of the cross-correlation $\langle \omega_{ij}^{o}(\theta_k)\omega_{ij}^{o}(\theta_l)\rangle$ has two contributions. We take field 'a' of the Millennium Simulation and two redshift bins as described in Section 2.4.1. The LEFT panel shows the clustering term calculated as described in Van Waerbeke (2010). The matrices are 6×6 and increase in scale from bottom to top and left to right. The RIGHT panel is the total covariance which includes the bootstrap covariance matrix in addition to the clustering covariance. The scale on the right is in units of 10^{-4}. Note that the clustering term results in a very flat covariance between all scales, whereas the bootstrap covariance adds relatively little and only on the smallest scales.

functions. These angular correlation functions can then be used to calculate the covariance.

This procedure suffices for all contributions to the covariance matrix Equation (2.26), except the first term $\langle \omega_{ij}^{o}(\theta_k)\omega_{ij}^{o}(\theta_l)\rangle$. As shown in Van Waerbeke (2010), the covariance of the cross-correlation function has a contribution due to the intrinsic clustering of the background and foreground populations. This so-called clustering term can be calculated analytically given the auto-correlation functions and the survey geometry. We add the covariance due to the clustering term to our bootstrap covariance matrix. The left panel in Figure 2.12 shows the covariance from the clustering term and the right panel shows the total covariance of the cross-correlation function between two redshift bins of the Millennium Simulation (see Section 2.4.1). This constitutes the largest contribution to the covariance matrix of Equation (2.26).

2.7 Solving the three-bin case analytically

For three bins it is easy to derive equations for the three observed cross-correlation functions, analogous to what is presented in Equation (2.14) for the two-bin case. Let the n×n matrix in Equation (2.11) be called F. The inverse can be calculated using the adjoint,

$$\mathrm{F}^{-1} = \frac{\mathrm{adj(F)}}{\det(\mathrm{F})}. \tag{2.27}$$

Both the adjoint (adj(F)) and the determinant (det(F)) can be calculated easily:

$$\begin{aligned}
\mathrm{adj(F)} &= \begin{pmatrix} f_{22}^2 f_{33}^2 - f_{32}^2 f_{23}^2 & f_{31}^2 f_{23}^2 - f_{21}^2 f_{33}^2 & f_{21}^2 f_{32}^2 - f_{31}^2 f_{22}^2 \\ f_{32}^2 f_{13}^2 - f_{12}^2 f_{33}^2 & f_{11}^2 f_{33}^2 - f_{31}^2 f_{13}^2 & f_{31}^2 f_{12}^2 - f_{11}^2 f_{32}^2 \\ f_{12}^2 f_{23}^2 - f_{22}^2 f_{13}^2 & f_{21}^2 f_{13}^2 - f_{11}^2 f_{23}^2 & f_{11}^2 f_{22}^2 - f_{21}^2 f_{12}^2 \end{pmatrix} \\
\mathrm{det(F)} &= f_{11}^2 f_{22}^2 f_{33}^2 + f_{21}^2 f_{32}^2 f_{13}^2 + f_{31}^2 f_{12}^2 f_{23}^2 \\
&\quad - f_{11}^2 f_{32}^2 f_{23}^2 - f_{21}^2 f_{12}^2 f_{33}^2 - f_{31}^2 f_{22}^2 f_{13}^2
\end{aligned} \tag{2.28}$$

With the inverse of of the matrix in hand we can now use Equation (2.11) to write the true auto-correlations in terms of the observed auto-correlations,

$$\omega_{\mathrm{ii}}^{\mathrm{T}} \mathrm{det(F)} = \omega_{\mathrm{ii}}^{\mathrm{o}} \left(\frac{N_{\mathrm{i}}^{\mathrm{o}}}{N_{\mathrm{i}}^{\mathrm{T}}}\right)^2 (f_{\mathrm{jj}}^2 f_{\mathrm{kk}}^2 - f_{\mathrm{kj}}^2 f_{\mathrm{jk}}^2) + \omega_{\mathrm{jj}}^{\mathrm{o}} \left(\frac{N_{\mathrm{j}}^{\mathrm{o}}}{N_{\mathrm{i}}^{\mathrm{T}}}\right)^2 (f_{\mathrm{ki}}^2 f_{\mathrm{jk}}^2 - f_{\mathrm{ji}}^2 f_{\mathrm{kk}}^2) + \omega_{\mathrm{kk}}^{\mathrm{o}} \left(\frac{N_{\mathrm{k}}^{\mathrm{o}}}{N_{\mathrm{i}}^{\mathrm{T}}}\right)^2 (f_{\mathrm{ji}}^2 f_{\mathrm{kj}}^2 - f_{\mathrm{ki}}^2 f_{\mathrm{jj}}^2), \tag{2.29}$$

where i≠j≠k. The three true auto-correlation functions are found by permutations of the indices (i,j,k)=(1,2,3), (2,3,1) and (3,1,2). Note that the equation is symmetric in the last two indices yielding the same result for (i,j,k) and (i,k,j). Substituting these equations into Equation (2.10) for the observed cross-correlations we find,

$$\begin{aligned}
\omega_{\mathrm{ij}}^{\mathrm{o}} \mathrm{det(F)} &= \omega_{\mathrm{ii}}^{\mathrm{o}} \frac{N_{\mathrm{i}}^{\mathrm{o}}}{N_{\mathrm{j}}^{\mathrm{o}}} \left[f_{\mathrm{ii}} f_{\mathrm{ij}} (f_{\mathrm{jj}}^2 f_{\mathrm{kk}}^2 - f_{\mathrm{kj}}^2 f_{\mathrm{jk}}^2) + f_{\mathrm{ji}} f_{\mathrm{jj}} (f_{\mathrm{kj}}^2 f_{\mathrm{ik}}^2 - f_{\mathrm{ij}}^2 f_{\mathrm{kk}}^2) + f_{\mathrm{ki}} f_{\mathrm{kj}} (f_{\mathrm{ij}}^2 f_{\mathrm{jk}}^2 - f_{\mathrm{jj}}^2 f_{\mathrm{ik}}^2) \right] \\
&+ \omega_{\mathrm{jj}}^{\mathrm{o}} \frac{N_{\mathrm{j}}^{\mathrm{o}}}{N_{\mathrm{i}}^{\mathrm{o}}} \left[f_{\mathrm{ii}} f_{\mathrm{ij}} (f_{\mathrm{ki}}^2 f_{\mathrm{jk}}^2 - f_{\mathrm{ji}}^2 f_{\mathrm{kk}}^2) + f_{\mathrm{ji}} f_{\mathrm{kk}} (f_{\mathrm{ii}}^2 f_{\mathrm{kk}}^2 - f_{\mathrm{ki}}^2 f_{\mathrm{ik}}^2) + f_{\mathrm{ki}} f_{\mathrm{kj}} (f_{\mathrm{ji}}^2 f_{\mathrm{ik}}^2 - f_{\mathrm{ii}}^2 f_{\mathrm{jk}}^2) \right] \\
&+ \omega_{\mathrm{kk}}^{\mathrm{o}} \frac{N_{\mathrm{k}}^{\mathrm{o}^2}}{N_{\mathrm{i}}^{\mathrm{o}} N_{\mathrm{j}}^{\mathrm{o}}} \left[f_{\mathrm{ii}} f_{\mathrm{ij}} (f_{\mathrm{ji}}^2 f_{\mathrm{kj}}^2 - f_{\mathrm{ki}}^2 f_{\mathrm{jj}}^2) + f_{\mathrm{ji}} f_{\mathrm{jj}} (f_{\mathrm{ki}}^2 f_{\mathrm{ij}}^2 - f_{\mathrm{ii}}^2 f_{\mathrm{kj}}^2) + f_{\mathrm{ki}} f_{\mathrm{kj}} (f_{\mathrm{ii}}^2 f_{\mathrm{jj}}^2 - f_{\mathrm{ji}}^2 f_{\mathrm{ij}}^2) \right]
\end{aligned} \tag{2.30}$$

Permuting the indices as above yields equations for the three observed cross-correlation functions. There are three equations and six unknowns –note that f_{ii}, f_{jj} and f_{kk} depend only on f_{ij}, f_{ik}, f_{ji}, f_{jk}, f_{ki} and f_{kj}. By considering more than one scale we can double the number of equations making the system constrained.

2.8 Conclusion and discussion

We have presented an analytic framework for estimating contamination between photometric redshift bins, without the need for any spectroscopic data beyond those used to train the photometric redshift code. To measure the contamination between redshift bins we exploit the fact that mixing between bins will result in a non-zero angular cross-correlation between those bins. We have shown how the contamination will affect the observed angular correlation functions for the general case of contamination between an arbitrary number of bins. For the case of two- and three-bins we explicitly work out the equations.

The case of two-bins is given special attention since it is the simplest case. We note that if the contamination between bins is small enough, then each pair of bins can be considered independently,

yielding an accurate measure of contamination between all bins. We refer to this as a global pairwise analysis.

We test our formalism with mock galaxy catalogues created from the Millennium Simulation. We verify that there is no evidence of contamination, finding an upper limit of $\sim$2 per cent at the 99.9 per cent confidence level. The catalogues are then contaminated by moving galaxies between redshift bins. We demonstrate that the two-bin analysis is able to recover input contamination between redshift bins. The effects of galaxy density and bin-width are investigated. We find that our ability to constrain the contamination fractions is not very sensitive to object density, whereas narrower bins offer better constraints.

We split the mock catalogues into four redshift bins and apply artificial contamination between all pairs. The global pairwise analysis is used to constrain the contamination fractions between all pairs of redshift bins. A Monte-Carlo method is then used to estimate the true (uncontaminated) redshift distribution, and the true average redshift of galaxies in each bin. This is valuable information for the cosmological interpretation of galactic surveys, and in particular weak lensing by large scale structure. We demonstrate the ability of the method to accurately recover the input contamination as well as reconstruct the true redshift distribution and average redshift of each bin.

The formalism is applied to a real galaxy survey; the four square degree deep component of the Canada-France-Hawaii Telescope Legacy Survey for which there are photometric redshift catalogues (Ilbert et al., 2006). We divide the data into four redshift bins and apply the global pairwise analysis. This yields constraints on the contamination fractions, the true redshift distribution, and the true average redshift of galaxies in each bin. We demonstrate here the feasibility of the method with only four square degrees of sky coverage; future application to large galaxy surveys will significantly improve constraining power.

This work has focused on the application of the two-bin and global pairwise methods. For a small number of redshift bins, with sufficiently small contamination, the global pairwise analysis offers a quick and easy means of assessing the contamination between redshift bins. The benefit is largely computational, it is very fast to constrain a model with only two free parameters. A more sophisticated method (such as a Monte-Carlo-Markov-Chain (MCMC)) will be needed to implement the full multi-bin approach. With only three redshift bins there is a total of six free parameters which already renders the simple maximum-likelihood approach impractical. Although the full multi-bin approach will yield the most accurate results, for some applications the more simplistic pairwise analysis should suffice.

Future work will need to take into account the effect of weak lensing magnification, which causes an angular cross-correlation between background galaxies and foreground lenses. Since foreground lenses boost the magnitude of background galaxies there are more close pairs detected between these redshift slices then one would expect from random placement. This effect is well understood and can be easily modelled and accounted for (Scranton et al., 2005), however, since it depends on cosmology and the redshifts of the lens and background galaxies it cannot be removed in a model-independent way. Dust extinction of the background sources is also important, reducing the brightness of lensed galaxies (Ménard et al., 2009). This de-magnification is wavelength dependant, and in visible passbands

is comparable in magnitude to lensing magnification, therefore, it will need to be accounted for along with magnification.

The expected amplitude of the angular cross-correlation due to magnification is small. Using the CFHTLS-Deep fields, and a magnitude cut similar to that used in this work, Van Waerbeke (2010) find the amplitude of the angular cross-correlation between the redshift bins $z = [0.1, 0.6]$ and $z = [1.1, 1.4]$ to be about 0.01 on scales smaller than 1 arcmin. While this is clearly an important contribution to the angular cross-correlations measured here it can only account for about 10 per cent of the observed signal on these scales. This is similar to the uncertainty introduced by the pairwise approximation, and within the error budget of this analysis.

Photometric redshifts are better measured for red galaxy types. Furthermore red and blue galaxies cluster differently, resulting in distinct angular correlation functions. Therefore, it is likely the case that galaxies doing the contamination are predominantly blue and exhibit a systematically different angular correlation than the red galaxies which do not contaminate. More work is needed to understand the severity of this bias. However, since any cross-correlation signal (above that expected from magnification) indicates contamination it is always possible to use this technique as a null test.

We have presented a method of measuring contamination between photometric redshift bins using the angular correlation function, and without any need for spectroscopically determined redshifts. The method is able to constrain the true redshift distribution and the true average redshift in a photometric bin, both of which are of keen interest to cosmological use of these data. The accuracy of this method will need to be improved to address the needs of high precision cosmology. The inclusion of the galaxy-shear correlation function to break parameter degeneracies has been investigated by Zhang et al. (2009), showing that the stringent requirements of future surveys can be reached if this information is included. Without the need for accurate weak lensing shear measurements, the method we present here is more accessible and provides valuable information.

Chapter 3

CFHTLenS tomographic weak lensing: quantifying accurate redshift distributions[1]

3.1 Introduction

Weak gravitational lensing by large-scale structure provides valuable cosmological information that can be obtained by analysing the apparent shapes of distant galaxies that have been coherently distorted by foreground mass (Bartelmann & Schneider, 2001). Since weak lensing is sensitive to the distance-redshift relation and the time-dependent growth of structure, it is a particularly useful tool for constraining models of dark energy (Albrecht et al., 2009, 2006; Peacock et al., 2006). To measure the contribution of dark energy over time, the lensing signal must be measured at several redshifts, this is known as weak lensing tomography (see for example, Hu, 1999; Huterer, 2002). Several observations of weak lensing tomography have been completed (Bacon et al., 2005; Massey et al., 2007; Semboloni et al., 2006). Most recently a study of the Cosmic Evolution Survey (COSMOS) by Schrabback et al. (2010) found evidence for the accelerated expansion of the Universe from weak lensing tomography.

Redshift information is vital to weak lensing interpretation since the distortion of light bundles is a geometric effect and the growth of structure is redshift-dependent. Weak lensing data sets necessitate the use of photometric redshifts due to the large number of galaxies they contain. Spectroscopic redshifts typically exist for a small and relatively-bright fraction of galaxies, providing a training set for photometric redshifts at brighter magnitudes. Several approaches for determining the redshift distribution of galaxies have been used in past weak lensing studies. Many early studies (see for example, Bacon et al., 2003; Benjamin et al., 2007; Fu et al., 2008; Hamana et al., 2003; Hoekstra et al., 2006; Jarvis et al., 2003; Van Waerbeke et al., 2005, 2002), lacking multi-band photometry, relied on external

[1] A version of this chapter has been published. J. Benjamin, L. Van Waerbeke, C. Heymans, M. Kilbinger, T. Erben, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, B. Rowe, T. Schrabback, F. Simpson, J. Coupon, L. Fu, J. Harnois-Déraps, M. J. Hudson, K. Kuijken, E. Semboloni, S. Vafaei, and M. Velander. CFHTLenS tomographic weak lensing: Quantifying accurate redshift distributions, 2013, MNRAS, Volume 431, Issue 2, pp.1547-1564

photometric redshift samples such as the Hubble Deep Field North and South, and the CFHTLS-Deep fields. Due to the small area of these fields, sampling variance was an important, but often neglected, source of error in these studies, as presented by Van Waerbeke et al. (2006).

Current and planned weak lensing surveys have multi-band photometry enabling photometric redshift estimates for all galaxies. Methods for measuring photometric redshifts use various model-fitting techniques with the goal of finding a match between the observed photometry and template galaxy spectra which are displaced in redshift and convolved with the optical response of the filter set, telescope, and camera. Depending on the set of photometric filters, degeneracies can exist between different template spectra at different redshifts. We refer to large errors in the best-fitting parameters due to mismatches under these degeneracies as catastrophic errors. The effect of catastrophic errors on weak lensing parameter constraints has been investigated in several studies, for example Bernstein & Huterer (2010); Ma et al. (2006) and Hearin et al. (2010). Using a detailed Fisher matrix analysis, Hearin et al. (2010) show the importance of properly characterising catastrophic errors to dark energy parameter constraints using weak lensing tomography. The implication of neglecting these errors is not well known, although Hearin et al. (2010) argue that there are many factors governing the final impact on dark energy parameters and each survey needs to be carefully considered to make any definitive statement. It is clear that catastrophic errors will become increasingly important in the next generation of weak lensing cosmic shear surveys.

In this chapter we present a tomographic weak lensing analysis of the Canada-France-Hawaii Telescope Lensing Survey[2] (CFHTLenS), with redshifts measured in Hildebrandt et al. (2012) using the Bayesian photometric redshift code (BPZ, Benítez, 2000). The BPZ analysis of the CFHTLenS photometry uses a set of 6 recalibrated spectral energy distribution galaxy templates from Capak et al. (2004) and a magnitude dependent prior on the redshift distribution (see Hildebrandt et al., 2012, for further details). If the galaxy template set and prior used are an accurate and complete representation of the true galaxy population at all redshifts, then the probability distribution function (PDF) calculated using BPZ determines the true error distribution. The redshift distribution of a galaxy sample can then be calculated from the sum of the PDFs to determine an accurate redshift distribution that includes the effects of both statistical and catastrophic errors. This is in contrast to the standard method of using a histogram of photometric redshifts taken from the maximum of the posterior. We test the accuracy of the summed PDFs with overlapping spectroscopic redshifts at bright magnitudes and with resampled COSMOS-30 redshifts (Ilbert et al., 2009) at faint magnitudes. In both cases we also assess the level of contamination between redshift bins using the angular cross-correlation technique presented in Chapter 2.

Demonstrating the accuracy of the summed PDFs is of particular interest to those using the CFHTLenS data products. However, the results in this chapter also contribute to a broader understanding of the use of redshift PDFs as measures of the redshift distribution. The CFHTLenS photometric redshifts are measured using BPZ (Benítez, 2000), however, other methods also produce posterior probability distribution functions such as the Zurich Extragalactic Bayesian Redshift Analyzer

[2]http://www.cfhtlens.org

(ZEBRA) photometric redshift code (Feldmann et al., 2006) and a photometric redshift-independent method presented in Lima et al. (2008) and extended in Cunha et al. (2009). The Cunha et al. (2009) method estimates a probability distribution based on the redshifts of nearest neighbour galaxies in multi-dimensional phase-space. There have been several studies focused on using redshift PDFs as estimates of the redshift distribution. Brodwin et al. (2006) show that the summed PDFs of galaxies can be used instead of the maximum likelihood values, as a better estimate of the redshift distribution of galaxies. This was tested on simulated galaxy samples using a Monte Carlo technique. Using the full PDF for galaxies has been shown to dramatically reduce the weak lensing calibration bias for galaxy-galaxy lensing (Mandelbaum et al., 2008). Wittman (2009) presents a technique to estimate an unbiased redshift using the PDF of a galaxy, effectively correcting the maximum likelihood value. The work in this chapter builds on these results and most notably we will use two methods to assess the accuracy of the summed PDFs at faint magnitudes where spectroscopic redshift coverage does not exist.

Previous CFHT Legacy Survey (CFHTLS) results were found to be biased, underestimating the shear at high redshifts, and requiring the addition of a nuisance parameter when model fitting (Kilbinger et al., 2009). Furthermore, the field selection excluded fields based on cosmology dependent criterion which possibly biased cosmological constraints. The CFHTLenS catalogues we use in this chapter have been thoroughly tested for systematic errors. These tests are cosmology insensitive and were completed without any cosmological analysis of the data (Heymans et al., 2012b). One of the primary goals of this chapter is to demonstrate that the redshift scaling of the shear is consistent with expectations. We limit our cosmic shear analysis to two broad redshift bins in order to obtain parameter constraints that do not depend on the modelling of intrinsic alignment (Croft & Metzler, 2000; Heavens et al., 2000; Hirata & Seljak, 2004b). A study of cosmological constraints from CFHTLenS with several redshift bins, accounting for intrinsic alignment is presented in Heymans et al. (2012a). Kilbinger et al. (2013) present a thorough investigation of 2D cosmic shear, including a comparison of all popular second order shear statistics. Simpson et al. (2013) use the tomographic shear signal presented in this chapter to constrain deviations from General Relativity on cosmological scales.

CFHTLenS has an effective area of 154 square degrees with deep photometry in five broad bands u^*, g', r', i', and z' and a 5σ point source limiting magnitude in the i'-band of $i'_{\rm AB}\sim25.5$. These data were obtained as part of the CFHTLS, which completed observations in early 2009. Details of the fields are presented in Figure 3.1. Heymans et al. (2012b) present an overview of the CFHTLenS analysis pipeline summarizing the weak lensing data processing with THELI (Erben et al., 2012), shear measurement with *lens*fit (Miller et al., 2013) and photometric redshift measurement from PSF-matched photometry (Hildebrandt et al., 2012) using BPZ. Each galaxy in the CFHTLenS catalogue has a shear measurement $\epsilon_{\rm obs}$, an inverse variance weight w, a PDF giving the posterior probability as a function of redshift, and a photometric redshift estimate from the peak of the PDF $z_{\rm p}$. The shear calibration corrections described in Miller et al. (2013) and Heymans et al. (2012b) are applied and we limit our analysis to the 129 of 171 pointings that have been verified as having no significant systematic errors through a series of cosmology-insensitive systematic tests described in Heymans et al. (2012b).

This chapter is organized as follows, in Section 3.2, we use a series of tests to determine whether

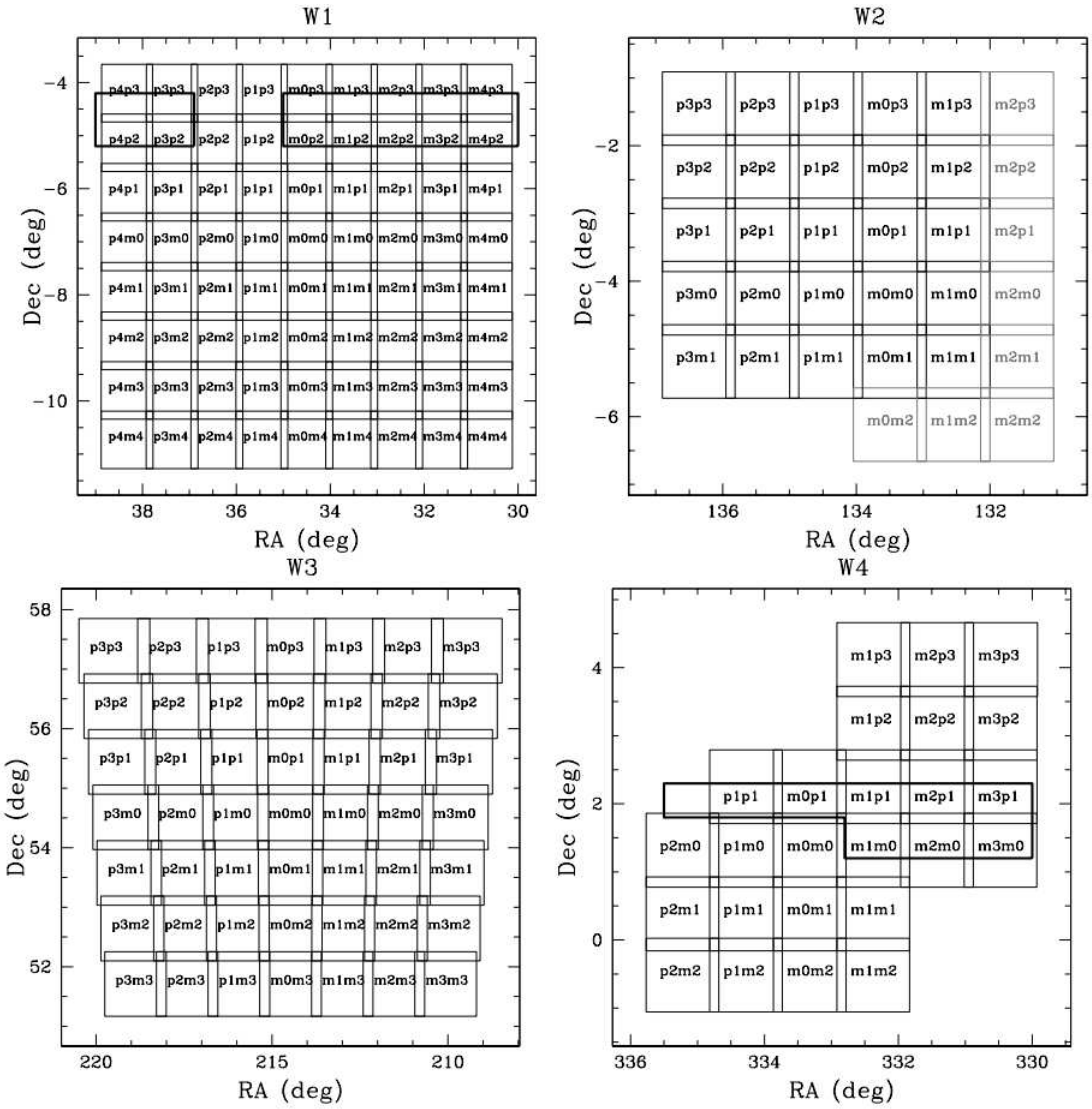

Figure 3.1: CFHTLenS is composed of the four wide fields from CFHTLS labelled W1, W2, W3, and W4. The number of pointings in each field varies with 72 in W1, 33 in W2, 49 in W3, and 25 in W4 for a total of 179 pointings. Full photometry was not obtained for the 8 grey pointings in W2 and they are not used in CFHTLenS. Overlap between fields and masking (Erben et al., 2012) brings the effective area of the remaining 171 pointings to 154 square degrees. The enclosed areas in W1 and W4 denote regions with spectroscopic coverage (Hildebrandt et al., 2012).

the PDFs are sufficiently accurate to determine the redshift distributions for the many different science analyses of the CFHTLenS data set, and then apply our findings to the first tomographic analysis of the CFHTLenS data set in Section 3.3. We investigate the effect of non-linear modelling of the mass power spectrum and baryons on our tomographic weak lensing results in Section 3.4. Section 3.5 contains our concluding remarks.

3.2 Galaxy redshift distributions determined from the photometric redshift PDF

When considering the redshift of an individual galaxy, a best-fitting redshift is often measured from the peak of the PDF (e.g. Hildebrandt et al., 2012; Ilbert et al., 2006). If many galaxies are considered, the sum of their PDFs can be used as an estimate of the redshift distribution instead of the distribution of best-fitting redshifts. We show in this section that, by using information from the entire PDF, we achieve an accurate model of the redshift distribution. The accuracy of the PDFs is not known *a priori* since this depends on whether the template spectral energy distributions and prior information are a representative and complete description of the galaxies in the survey.

We compare the summed PDFs against several other methods of measuring the redshift distribution. These methods include a comparison with the overlapping VVDS and DEEP2 spectroscopic redshifts (see Section 3.2.1), statistical resampling of the CFHTLenS photometric redshifts using the COSMOS-30 redshifts (see Section 3.2.2), and a photometric redshift contamination analysis (see Section 3.2.3).

We divide the data into six redshift bins and measure the redshift distribution of each. We are limited in the total number of bins by the pairwise contamination analysis, which breaks down for larger numbers of bins (see Section 3.2.3 for a more detailed discussion). The redshifts are most reliable in the range $0.1 < z_\mathrm{p} < 1.3$ where comparison to spectroscopic redshifts, for $i' < 24.5$, shows the scatter to be $0.03 < \sigma_{\Delta z} < 0.06$, with an outlier rate of less than 10 per cent (Hildebrandt et al., 2012). Here $\sigma^2_{\Delta z}$ is the variance in the value of Δz, which is given by

$$\Delta z = \frac{z_\mathrm{p} - z_\mathrm{s}}{1 + z_\mathrm{s}}, \tag{3.1}$$

where z_p and z_s are the photometric and spectroscopic redshifts, respectively.

The redshift bins are chosen such that each bin is approximately four times wider than the photometric redshift error $0.04(1+z_\mathrm{p})$. This is done to avoid excessive contamination between adjacent bins. Since the high confidence photometric redshift range ends at $z = 1.3$ we ensure this is a bin edge. For $z_\mathrm{p} > 1.3$ there are only a small number of galaxies, making a subdivision of this range difficult. Similarly, a lack of objects at low redshift prevents us from selecting $z = 0.1$ as a bin edge, despite this being the boundary of the high confidence region. The six bins are as follows:

Bin 1: $0.00 < z_1 \leq 0.17$
Bin 2: $0.17 < z_2 \leq 0.38$
Bin 3: $0.38 < z_3 \leq 0.62$
Bin 4: $0.62 < z_4 \leq 0.90$

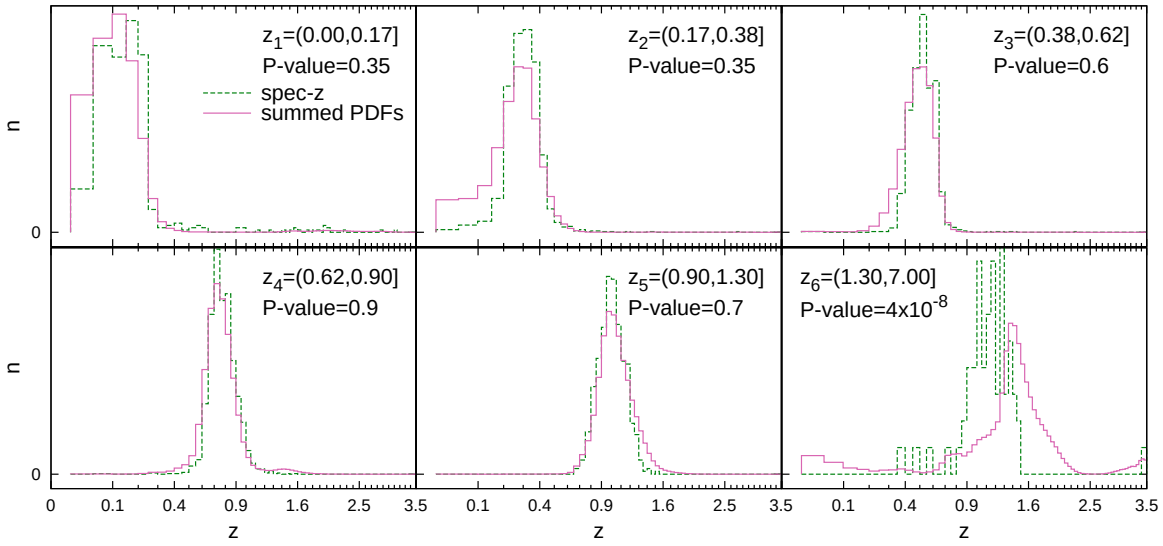

Figure 3.2: Comparison of the predicted redshift distributions within each broad photometric redshift bin, labelled $z_{\rm i}$. A magnitude cut of $i' < 23.0$ is used for comparison with spectroscopic redshifts. Solid lines (pink) show the summed PDFs for all galaxies within a given photometric redshift bin. Dashed lines (green) show the spectroscopic redshift distribution. The listed P-values are the result of a two-sample Kolmogorov-Smirnov test of the distributions, we adopt a significance level of $\alpha = 0.05$ rejecting the null hypothesis that the two distributions are drawn from the same population for the highest redshift bin.

Bin 5: $0.90 < z_5 \leq 1.30$
Bin 6: $1.30 < z_6$

3.2.1 Comparison with spectroscopic redshifts

We begin by investigating the redshift distribution given by spectroscopic redshifts. Spectroscopic redshifts from the VIMOS[3] VLT[4] Deep Survey (VVDS, Le Fèvre et al., 2005) and the DEEP2 galaxy redshift survey (Newman et al., 2012) overlap with CFHTLenS and were used to test the photometric redshifts. For a given photometric redshift bin we can select those galaxies that have spectroscopic redshifts and examine their redshift distribution. The spectroscopic sample is complete for $i' \lesssim 22.0$, dropping to $\sim$90 per cent completeness for $i' < 23.0$. We adopt the latter cut to ensure that there are a sufficient number of galaxies for our analysis. The catalogues are also cut to exclude objects on masked regions and those that are flagged as stars. Stars are selected with `star_flag` (see Erben et al., 2012, for more details). Due to the dithering pattern, which ensures that exposures exist between individual CCD chips, there is a variable number of exposures over a single pointing. This changing photometric depth is difficult to account for when constructing a random catalogue with the same properties, which is necessary for the contamination analysis presented in Section 3.2.3. To avoid this complexity, a final cut is made to select galaxies on areas of the sky that were detected during every exposure, and random

[3]VIsible imaging Multi-Object Spectrograph
[4]Very Large Telescope array

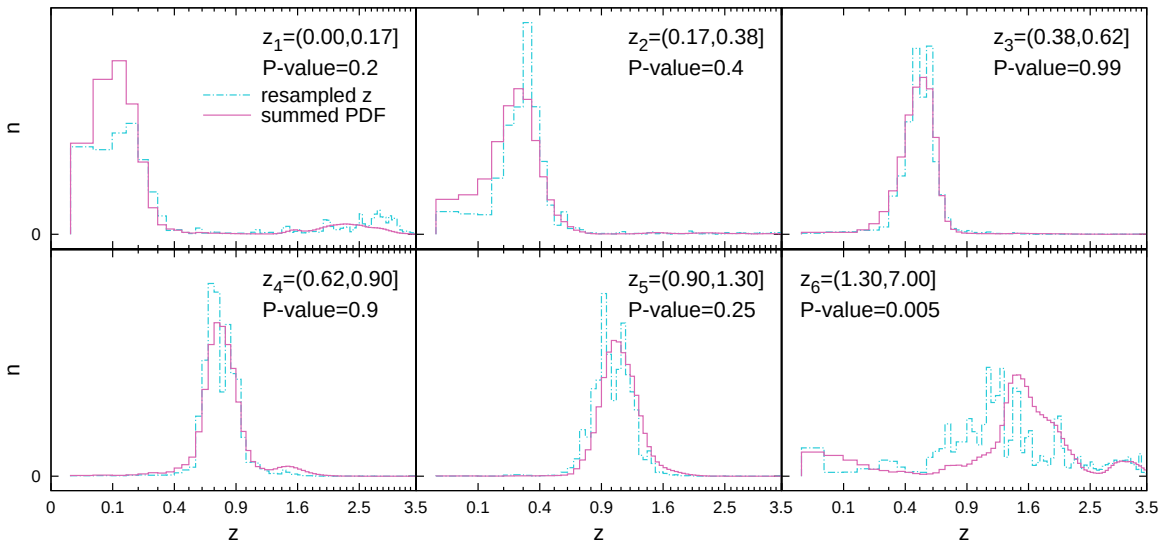

Figure 3.3: Comparison of the predicted redshift distributions with a magnitude cut of $i' < 24.7$. Solid lines (pink) show the summed PDFs for all galaxies within a given photometric redshift bin. Dot-dashed histogram (cyan) shows the result of resampling the CFHTLenS redshifts using the constructed conditional probability $P(z_{30}|z_{\rm p})$. The P-values are the result of a KS test, we reject the null hypothesis for the highest redshift bin at $\alpha = 0.05$.

objects are placed only in these areas. We do not expect this to bias our results as there is no correlation between the physical properties of a galaxy and which part of the CCD mosaic it was observed on.

A comparison of redshift distributions for $i' < 23.0$ is presented in Figure 3.2. For each redshift bin we show the redshift distribution predicted by the summed PDF (solid line), and the spectroscopic redshift distribution (dashed line). The PDFs of all galaxies within a given redshift bin are summed and the resulting distribution normalised to obtain the solid line. If the summed PDF is a good representation of the true error distribution, then we would expect this distribution to agree with the redshift distribution measured with the spectroscopic redshifts.

We use the Kolmogorov-Smirnov two-sample test (KS test) to determine if the two distributions in Figure 3.2 are consistent with being drawn from the same population (details of this test can be found in, for example, Wall & Jenkins, 2003). Before performing the test we adopt as a discriminating criterion a significance level of $\alpha = 0.05$. The P-values found from the KS test are presented as labels in Figure 3.2. We find that the distributions for the first five redshift bins are consistent with having been drawn from the same population at a significance level of $\alpha = 0.05$. However, we can reject the null hypothesis for the last redshift bin at the same level of significance, indicating that the two distributions are significantly different. This is indicative of the large uncertainties in the photometric redshifts at $z_{\rm p} > 1.3$ and confirms the choice of this cut-off made by Hildebrandt et al. (2012).

3.2.2 Comparison with COSMOS photometric redshifts

The agreement at bright magnitudes shown in Figure 3.2 is encouraging; however, the majority of lensing studies include fainter galaxies, for example a magnitude limit of $i' < 24.7$ is adopted for the

measurement of CFHTLenS galaxy shapes (Miller et al., 2013). Therefore, we wish to investigate the redshift distribution with this deeper magnitude cut. The spectroscopic redshift sample described in Section 3.2.1 cannot be used for comparison since the completeness of this sample drops sharply beyond $i' \sim 23$. Instead, we use the COSMOS-30 photometric redshift catalogue (Ilbert et al., 2009), which is accurate to $\sigma_{\Delta z} \simeq 0.012$ due to 30 bands of wavelength coverage from the ultraviolet to the mid-infrared. Quoting values for the Subaru *i*-band, the COSMOS-30 data are 99.8 percent complete for $i < 25.5$, and have a 5σ point source limiting magnitude of $i \sim 26.2$ (Ilbert et al., 2009). A resampling procedure is used to estimate the redshift distribution of deep CFHTLenS galaxies based on the distribution of COSMOS-30 redshifts with CFHTLS overlap.

Although the 1.6 square degree COSMOS field contains the one square degree CFHTLS-Deep field D2, there are no overlapping CFHTLS-Wide fields. Therefore, it is not possible to directly match CFHTLenS galaxies to objects in the COSMOS-30 catalogue. This issue can be circumvented in a novel way using the photometric catalogue of D2 provided in Hildebrandt et al. (2009). Using the fact that the photometric systems for the CFHTLS-Wide and Deep data are identical we add random Gaussian noise, scaled to simulate data taken at CFHTLS-Wide depth, to the magnitude estimates in the D2 photometric catalogue. Using artificially-degraded catalogues generated in this way, we calculate a Wide-like photometric redshift estimate $z_{\rm p}$ using the maximum of the posterior distribution as described in Hildebrandt et al. (2012). This is done for each D2 object in a catalogue matched to the COSMOS-30 catalogue of Ilbert et al. (2009), employing an association radius of 1.0 arcsecond. The COSMOS-30 redshifts in the matched catalogue are labelled z_{30}.

This matched catalogue of z_{30} and noise-degraded Wide-like $z_{\rm p}$ estimates can then be used to acquire information about the joint probability distribution of COSMOS-30 and CFHTLenS redshifts. We generate 100 realisations of the artificially degraded Wide-like catalogues, running the Bayesian photometric redshift estimation of Hildebrandt et al. (2012) for each realisation. Using this ensemble of $(z_{30}, z_{\rm p})$ pairs, we construct a two-dimensional histogram of galaxy number counts in square bins of width 0.0025 in redshift for both z_{30} and $z_{\rm p}$. This histogram is then used as an empirical estimate of the conditional probability density function $P(z_{30}|z_{\rm p})$ and allows us to estimate the corresponding cumulative probability distribution function $P(< z_{30}|z_{\rm p})$ for each $z_{\rm p}$ bin. Then, using inversion sampling from $P(< z_{30}|z_{\rm p})$ with a uniform pseudo-random number generator, samples of redshifts distributed according to $P(z_{30}|z_{\rm p})$ can be drawn.

With the assumption that $P(z_{\rm true}|z_{\rm p}) = P(z_{30}|z_{\rm p})$, the contamination in tomographic redshift bins can be estimated by resampling CFHTLenS redshifts according to $P(z_{30}|z_{\rm p})$. The resulting redshift distributions predicted from this method are given as dot-dash lines in Figure 3.3, and the summed PDFs are presented as solid lines. Note that the fine structure seen in the resampled redshifts is due to structures in the COSMOS field and does not represent real structures in the distribution of CFHTLenS galaxies. The small size of COSMOS means that it is limited by sample variance and individual clusters are able to leave an imprint in the resampled galaxies. This only affects the fine details of the resampled redshifts leading to a breakdown of the assumption $P(z_{\rm true}|z_{\rm p}) = P(z_{30}|z_{\rm p})$ for small redshift intervals.

We again adopt the null hypothesis that the two distributions are drawn from the same population.

Using a KS test we find that the null hypothesis can be rejected at a significance level of $\alpha = 0.05$ for the $z_p > 1.3$ redshift bin, but not for any of the other bins. The P-values found from the KS test are presented as labels in Figure 3.3. Our results again confirm that the CFHTLenS photometric redshifts of the $z_p > 1.3$ galaxies are unreliable. However, we find no evidence that the galaxies at $z_p < 0.1$, which is the lower limit for the high-confidence redshift range (Hildebrandt et al., 2012), are unreliable. This is likely because our lowest redshift bin extends to $z_p = 0.17$ and is therefore dominated by galaxies with well-measured photometric redshifts.

3.2.3 Redshift contamination from angular correlation functions

In order to further test the accuracy of the photometric redshift PDFs, we measure the redshift contamination using the angular cross-correlation technique detailed in Chapter 2. This method does not rely on many assumptions and is sensitive to any contamination between redshift bins. Since it only relies on the angular correlation function of the galaxies it is independent of the other methods used and serves as a critical test.

Overview of method

Galaxies cluster in over-dense regions, leading to an excess in the number of pairs found at a separation θ when compared to a random distribution of points. The two-point angular correlation function $\omega(\theta)$ quantifies this excess probability of finding pairs. A common estimator (Landy & Szalay, 1993) is

$$\omega_{\rm ij} = \frac{(D_{\rm i}D_{\rm j})_\theta}{(RR)_\theta}\frac{N_{\rm R}N_{\rm R}}{N_{\rm i}N_{\rm j}} - \frac{(D_{\rm i}R)_\theta}{(RR)_\theta}\frac{N_{\rm R}}{N_{\rm i}} - \frac{(D_{\rm j}R)_\theta}{(RR)_\theta}\frac{N_{\rm R}}{N_{\rm j}} + 1, \tag{3.2}$$

where $(D_{\rm i}D_{\rm j})_\theta$ is the number of pairs separated by a distance θ between data sets i and j, $(RR)_\theta$ is the number of pairs separated by a distance θ for a random set of points, $(D_{\rm i}R)_\theta$ is the number of pairs separated by a distance θ between data set i and a random set of points, $N_{\rm R}$ is the number of points in the random sample, and $N_{\rm i}$ ($N_{\rm j}$) is the number of points in data sample i (j). The auto-correlation is described by the case i=j, and the cross-correlation by the case i$\neq$j. Our analysis would hold for any estimator of the angular correlation function.

In the ΛCDM model, galaxies in well-separated non-overlapping redshift bins are not significantly clustered with each other. Therefore, clustering between these bins should be consistent with a random distribution of points, resulting in $\omega_{\rm ij} = 0$. Adjacent redshift bins will have a small positive $\omega_{\rm ij}$ owing to galaxy clustering at their shared edge, which becomes more pronounced for narrow redshift bins. If any non-zero angular cross-correlation is detected between the photometric redshift bins, they must share galaxies with similar redshifts.

As shown in Chapter 2, this simple realisation can be exploited to estimate contamination between photometric redshift bins. The reader is referred to that work for the full details of the method. Here we present only a few key equations and concepts before we apply the method to the CFHTLenS data.

The contamination fraction, $f_{\rm ij}$ is defined as the number of galaxies contaminating bin j from bin i as a fraction of the total number of galaxies $N_{\rm i}^{\rm T}$ which have a spectroscopic redshift that lies within redshift

bin i. If there is no overlap or contamination between redshift bins, $f_{ij} = 0$ when $i \neq j$, and $N_i^T = N_i^o$, where N_i^o is the total number of galaxies which have a photometric redshift that lies within redshift bin i. In the standard case of overlapping photometric redshift bins, the contamination fraction relates the observed number of galaxies in each photometric redshift bin N_i^o to the true underlying number of galaxies N_i^T as follows;

$$\begin{pmatrix} N_1^o \\ N_2^o \\ \cdots \\ N_m^o \end{pmatrix} = \begin{pmatrix} f_{11} & f_{21} & \cdots & f_{m1} \\ f_{12} & f_{22} & \cdots & f_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ f_{1m} & f_{2m} & \cdots & f_{mm} \end{pmatrix} \begin{pmatrix} N_1^T \\ N_2^T \\ \cdots \\ N_m^T \end{pmatrix}, \tag{3.3}$$

where m is the number of redshift bins and $f_{ii} = 1 - \sum_{k \neq i}^{m} f_{ik}$. We determine the contamination fractions f_{ij} from measurements of ω_{ij}^o, the observed two-point correlation function between photometric redshift bins i an j. For two redshift bins it can be shown that,

$$\omega_{12}^o = \frac{\omega_{11}^o \left(\frac{N_1^o}{N_2^o}\right) f_{12}(1 - f_{21}) + \omega_{22}^o \left(\frac{N_2^o}{N_1^o}\right) f_{21}(1 - f_{12})}{(1 - f_{12})(1 - f_{21}) + f_{12} f_{21}}. \tag{3.4}$$

Since ω_{12}^o is measured at multiple angular scales it is possible to determine the contamination fractions f_{12} and f_{21}. Degeneracy exists between f_{12} and f_{21}, which can be broken if the angular auto-correlation functions ω_{11}^o and ω_{22}^o have significantly different shapes. We expect this to be the case for the angular correlation functions of galaxy samples at different redshifts.

When considering more than two redshift bins we measure the contamination fractions using Equation (3.4) for each pair of bins in turn. This pairwise approximation assumes that higher-order contamination can be safely ignored, that is, the angular cross-correlation is not affected by the mutual contamination of the pair of bins by another redshift bin. As the number of bins increases or if the contamination fractions become large, this assumption is no longer valid and the method breaks down.

Once the contamination fractions f_{ij} have been measured, we can invert[5] the contamination matrix in Equation (3.3) to determine the true underlying number of galaxies in each redshift range N_i^T from the observed number of galaxies in each photometric redshift bin N_i^o. The true redshift distribution $n^i(z_j)$ for each photometric redshift bin i is then calculated over the full redshift range, sampled at each redshift z_j from

$$n^i(z_j) = f_{ij} N_i^T. \tag{3.5}$$

Contamination analysis

In order to estimate the contamination fractions in the CFHTLenS data the angular correlation functions must be measured. Brute-force pair counting algorithms are $\mathcal{O}(N^2)$, where N is the number of galaxies, which for our data-set results in prohibitively large computation times. The publicly available code ATHENA[6] employs a tree data structure to increase the speed of pair counting to $\mathcal{O}(Nlog(N))$ at the

[5]Since we expect the non-diagonal contamination fractions to be small the matrix should be diagonally dominant and therefore invertible.

[6]http://www2.iap.fr/users/kilbinge/athena/

cost of accuracy. The level of approximation is parametrized by the opening angle. Larger values indicate larger approximations with an opening angle of zero representing no approximation. Galaxies are grouped together into nodes in the tree data structure based on angular position. The structure is a hierarchy with the nodes on top containing more galaxies. The opening angle determines when to descend to lower nodes and higher spacial resolution. Tests of ATHENA against a more simplistic and robust algorithm are used to determine that with an opening angle of 0.03 we are making at most a one per cent error on the angular correlation function. This value of opening angle is used when measuring the angular correlation function.

For each pointing we measure the angular correlation function in six angular bins spaced logarithmically on the range $0.15 < \theta < 30$ arcmin. Above 30 arcmin the signal is very small providing little additional information. For each pointing the contamination fractions are estimated via the angular correlation function as outlined in Section 3.2.3. The covariance is estimated via a bootstrap technique, with an additional contribution coming from the field-to-field variance for the angular cross-correlations. The details of the maximum likelihood technique and covariance matrix are presented in Section 2.6. The likelihoods for the contamination fractions from each field are then combined with equal weighting.

The following matrices contain the measured contamination fractions with 68 per cent confidence regions. All values are multiplied by one hundred for ease of viewing. For the bright sample, $i' < 23.0$, we find,

$$f_{\rm ij} = \begin{pmatrix} 65\pm5 & 4\pm1 & <1 & <1 & <1 & 6\pm6 \\ 28\pm4 & 87\pm3 & 8\pm2 & <1 & <1 & 7\pm7 \\ 1\pm1 & 7\pm2 & 85\pm3 & 9\pm2 & 1\pm1 & 7\pm7 \\ 1\pm1 & <1 & 6\pm2 & 85\pm2 & 38\pm6 & 5\pm5 \\ <1 & <1 & <1 & 4\pm1 & 56\pm7 & 29\pm12 \\ 3\pm1 & <1 & <1 & <1 & 3\pm1 & 18\pm16 \end{pmatrix}. \tag{3.6}$$

With a cut of $i' < 24.7$ we measure,

$$f_{\rm ij} = \begin{pmatrix} 42\pm8 & 3\pm2 & <1 & 1\pm1 & <1 & 9\pm6 \\ 32\pm4 & 75\pm3 & 8\pm2 & <1 & <1 & 1\pm1 \\ <1 & 18\pm2 & 79\pm3 & 7\pm2 & <1 & 1\pm1 \\ 4\pm4 & <1 & 11\pm2 & 78\pm3 & 20\pm3 & 3\pm3 \\ <1 & <1 & <1 & 9\pm2 & 73\pm4 & 42\pm5 \\ 18\pm4 & 2\pm1 & <1 & 4\pm1 & 5\pm3 & 36\pm6 \end{pmatrix}. \tag{3.7}$$

Many of the contamination fractions are one per cent deviations from zero, which is expected since we have this level of uncertainty in our estimation of the angular correlation functions. Note that the $\rm i^{th}$ column contains the location of all bin i galaxies. Due to the pairwise treatment of redshift bins columns do not sum to exactly 100 per cent. These matrices are extremely well conditioned with condition numbers of 8.72 and 6.35 for the $i' < 23.0$ and $i' < 24.7$ cases respectively, indicating that matrix inversion is numerically stable and does not contribute a significant uncertainty to the solution of Equation (3.3).

With the contamination fractions measured, the true number of galaxies in each redshift bin can

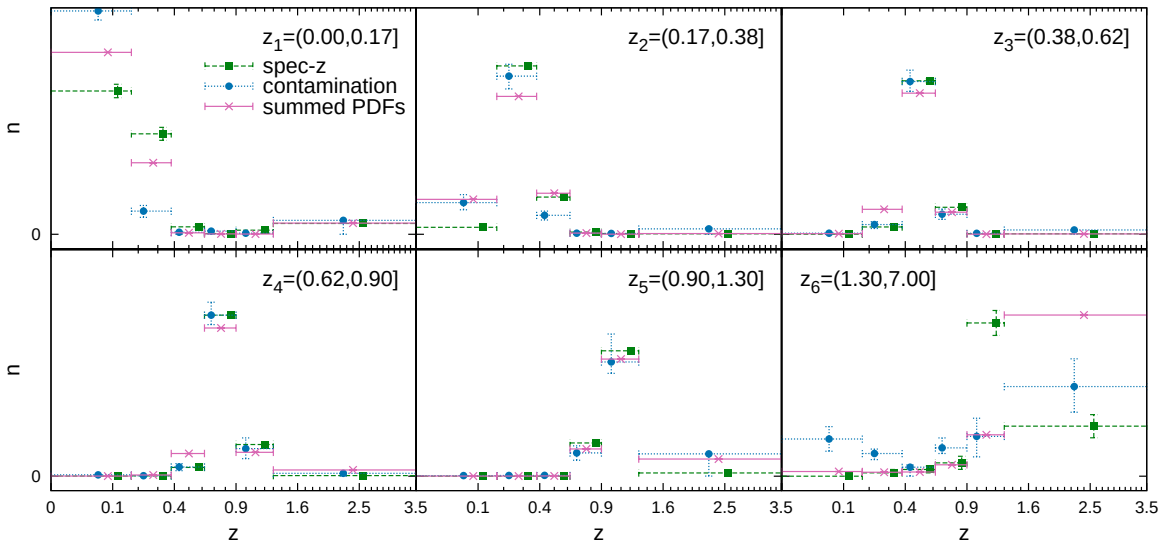

Figure 3.4: Comparison of the predicted true redshift distribution within each broad photometric redshift bin, labelled z_i. A magnitude cut of $i' < 23.0$ is used for comparison with spectroscopic redshifts. All horizontal error bars denote the width of the redshift bin and points are offset horizontally for clarity. Crosses with solid lines (pink) denote the summed PDFs when integrated within a given broad redshift bin, the error is calculated as the standard deviation from 1000 bootstrap samples. Filled circles with dotted lines (blue) show the result from our contamination analysis with 68 per cent confidence region. Filled squares with dashed lines (green) show the spectroscopic redshift data integrated within each broad redshift bin. The error is the standard deviation of 1000 bootstrap samples.

be calculated from Equation (3.3). The redshift distribution is then found from Equation (3.5). This is done with a Monte Carlo procedure for finding global solutions to the contamination matrix presented in Section 2.4.4. We can now compare our contamination results with those found from the spectroscopic redshifts and the COSMOS-30 photometric redshifts. However, since our contamination results exist in only six redshift bins we must also sum the distributions shown in Figures 3.2 and 3.3 within these six redshift bins.

For $i' < 23.0$ we present our contamination results in Figure 3.4. For each redshift bin we show the redshift distribution predicted by the PDFs (crosses with solid lines), the contamination analysis (filled circles with dotted lines), and the spectroscopic redshift distribution (filled squares with dashed lines). The horizontal error bars on all points denote the width of the redshift bins. The vertical scale is proportional to the number of galaxies but uses arbitrary units. For the summed PDF, the vertical error bar is calculated as the standard deviation of the summed PDFs for 1000 bootstraps of the galaxies within each bin z_i (this error bar is smaller than the points in Figure 3.4). Note that given the large number of galaxies in each bin ($\sim 40,000 - 600,000$) the statistical error of the summed PDF is very small. The vertical error bars on the contamination results enclose the 68 per cent confidence region which comes from a procedure for finding global solutions to the contamination (Section 2.4.4). The vertical error on the spectroscopic redshift distribution is taken as the standard deviation from 1000 bootstraps of the spectroscopic catalogue. For both cases where bootstraps are used we verified that

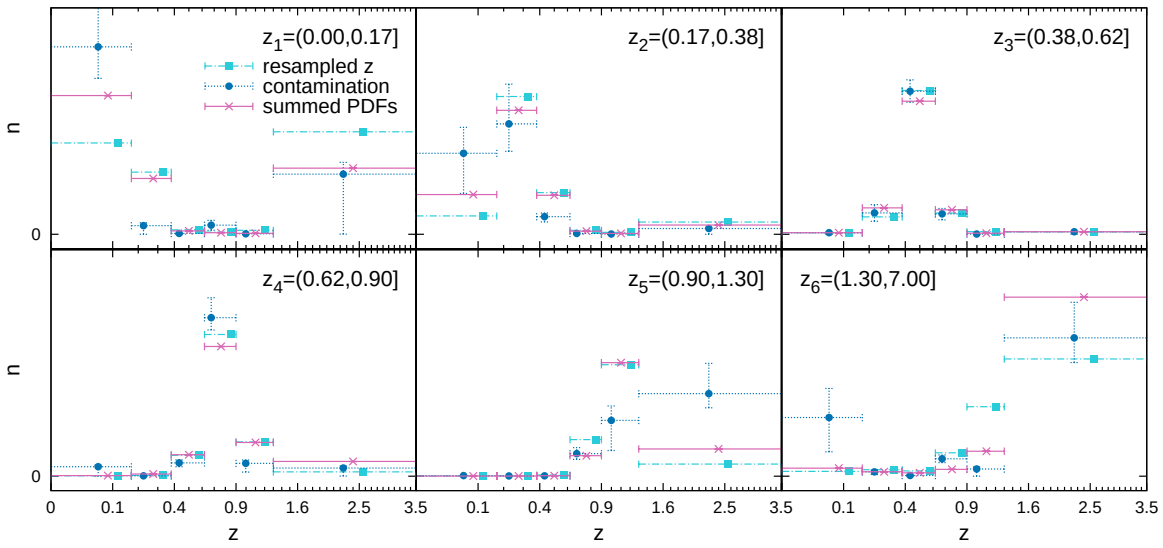

Figure 3.5: Same as Figure 3.4 except for the following differences. A magnitude cut of $i' < 24.7$ is used. Filled squares with dot-dashed line (cyan) show the resampled COSMOS-30 data integrated within each broad redshift bin. The error is given as the standard deviation of the 100 low-resolution reconstructions (see Section 3.2.2).

1000 bootstraps yields stable error estimates. For each bootstrap, objects are sampled with replacement and the resulting redshift distributions measured, the total number of galaxies sampled is equal to the number in the original catalogues.

Figure 3.4 shows the predicted redshift distribution for each of the six redshift bins used. The contamination points for a given sub-plot are contained within the corresponding row of the contamination matrix in Equation (3.6). For example, the top row shows that the majority of galaxies from bin 1 remain in bin 1, with $f_{11} = 65 \pm 5$ per cent. Contamination from other bins is less than the per cent level except for the neighbouring bin $f_{21} = 4 \pm 1$ per cent and the highest redshift bin $f_{61} = 6 \pm 6$. Keep in mind that the relative heights of points in the z_1 sub-plot do not follow these contamination values because the contamination $f_{\rm ij}$ represents the number of galaxies in bin j from bin i divided by the true number in bin i. However, investigating the matrix in relation to Figure 3.4 can help in grasping the presented information. We expect that the spectroscopic redshift distribution is the true distribution, assuming that the limited area of the spectroscopic samples does not bias the results, which is a reasonable assumption for our purposes. The contamination model is in poor agreement with the spectroscopic sample in the z_1 and z_6 sub-plots. For z_1 the contamination model underpredicts the contamination from bin 2 to bin 1, underestimating f_{21} as evidenced by the discrepancy between the contamination point and the spectroscopic point in the second bin of the z_1 sub-plot. Similarly the contamination is overpredicted for f_{12} which is seen in the first bin of the z_2 sub-plot. This represents a fundamental degeneracy in the angular cross-correlation method. Although an angular cross-correlation is detected between these two bins, unless the angular auto-correlations have significantly different slopes, the method cannot distinguish easily between bin 1 galaxies contaminating bin 2 or vice versa. A similar degeneracy explains the discrepancies in the z_6 sub-plot, there we see that f_{56} predicted by the angular cross-correlation is

too low and f_{65} in the z_5 sub-plot is too high. The contamination between these bins is detected but the direction of scatter is misidentified.

We use a KS test to determine if the distributions in Figure 3.4 are consistent with being drawn from the same population. Since there are three distributions and the KS test is a two-sample test we apply it to each pair of distributions. Furthermore, due to the small number of bins we must rely on tabulated critical values which exist for very few significance levels, therefore we are not able to list the P-values for each redshift bin. For each pair of distributions we find that we cannot reject the null hypothesis (drawn from the same population) at a significance level of $\alpha = 0.05$ for any of the redshift bins.

We present the results of the contamination analysis for $i' < 24.7$ in Figure 3.5. The summed PDF and contamination results are presented similarly to Figure 3.4. The resampled redshifts using COSMOS-30 are given as the dot-dash line and filled squares. The vertical error on the COSMOS-30 points is taken as the standard deviation of the 100 low-resolution resamplings (see Section 3.2.2). If we compare the contamination results to the resampled redshifts the greatest discrepancies are for the z_1, z_5, and z_6 sub-plots. The $f_{12} - f_{21}$ and $f_{56} - f_{65}$ degeneracies noted for the bright sample above appear again in Figure 3.5. Additionally the resampled redshifts predict a larger f_{61} and smaller f_{16} than do the contamination results which can be seen in the first and last bins of the z_1 and z_6 sub-plots respectively. The contamination analysis predicts that a significantly lower number of galaxies belong in bin 5 compared to the other methods, see bin 5 of the z_5 sub-plot. However, this is not due to scattering of bin 5 galaxies elsewhere (note that $f_{55} = 73 \pm 4$ per cent) instead the contamination analysis simply predicts fewer galaxies occupying this bin. To determine if these differences are statistically significant we use a KS test. We again have three distributions and apply the test between each pair. For each pair of distributions we find that we cannot reject the null hypothesis (drawn from the same population) at a significance level of $\alpha = 0.05$ for any of the redshift bins.

When using the finely binned spectroscopic and resampled redshifts in Sections 3.2.1 and 3.2.2 we were able to reject the null hypothesis for the high redshift bin $z_p > 1.3$. The smallest P-value found was in the high redshift bin when comparing the summed PDF with the spectroscopic distribution in Figure 3.2. Performing the same comparison with these distributions when summed within the six redshift bins of the contamination analysis we are not able reject the null hypothesis. The coarse binning required by the contamination analysis has reduced the statistical power of the test. However, the contamination analysis provides a complementary estimation of the redshift distribution, which agrees well with the other estimates and strengthens our confidence in the summed PDF as an accurate measure of the redshift distribution.

We conclude that the summed PDF can be used to estimate the redshift distribution for the high confidence redshift range $0.1 < z_p < 1.3$ determined by Hildebrandt et al. (2012). The comparison with the resampled COSMO-30 redshifts and the contamination analysis for $i' < 24.7$ suggest that an accurate estimate of the redshift distribution, including statistical and catastrophic errors, can be obtained from the sum of the PDFs. This result suggests that the model galaxy spectra and priors used in Hildebrandt et al. (2012) are a fair and sufficiently complete representation for the population of galaxies studied here.

3.3 Weak lensing tomography

In this chapter we present an analysis of the CFHTLenS tomographic weak lensing signal using two broad redshift bins and compare our results with a 2D analysis over the same redshift range. Setting our analysis in a flat ΛCDM cosmology framework, the initial aim is to use the consistent results we find between successive tomographic bins as a demonstration that the CFHTLenS catalogues are not subject to the redshift-dependent systematic biases that were uncovered in an earlier analysis of CFHTLS data (Kilbinger et al., 2009). This cosmology-dependent demonstration is the last in an extensive series of tests, which investigate the robustness and accuracy of the CFHTLenS catalogues. We stress, however, that this analysis was performed after the conclusion of a series of cosmology-insensitive tests presented in Heymans et al. (2012b) and the photometric redshift accuracy analysis presented in Section 3.2. Most importantly, no feedback loop existed between this cosmology-dependent test and the systematics and image simulation tests that determined the calibration corrections and the subset of reliable data that we use from the survey.

We choose to use two broad mid-to-high redshift bins for our tomographic analysis in order to reduce the potential contamination to the signal from intrinsic galaxy alignments (see, for example, Heavens et al., 2000; Heymans et al., 2012a, and references therein). We estimate the expected contamination of the measured weak lensing signal using the linear tidal field intrinsic alignment model of Hirata & Seljak (2004b), and following Bridle & King (2007b) by fixing its amplitude to the observational constraints obtained by Brown et al. (2002b). By limiting the redshift bins to photometric redshifts $0.5 < z_{\rm p} \leq 0.85$ and $0.85 < z_{\rm p} \leq 1.3$, we estimate that any contamination from intrinsic alignments is expected to be no more than a few per cent for each redshift bin combination. We therefore ignore any contributions from intrinsic alignments in this analysis as they are expected to be small in comparison to our statistical errors. Note that a low level of contamination would not be expected if we instead used the 6 narrow redshift bins that were analysed in the redshift contamination analysis in Section 3.2. We present a fine 6-bin tomographic analysis of the data in Heymans et al. (2012a) where the impact of intrinsic galaxy alignments is mitigated via the simultaneous fit of a cosmological model and an intrinsic alignment model. The findings of Heymans et al. (2012a) support the approach taken in this chapter to neglect the contribution of intrinsic alignments for our choice of redshift bins.

The 2D lensing analysis presented here is restricted to the same redshift range used in our tomographic analysis $0.5 < z_{\rm p} \leq 1.3$. We measure the shear correlation function on angular scales from $\sim$1 to $\sim$40 arcmin. The upper limit is set by our ability to measure the covariance matrix from simulations, see Section 3.3.3.

3.3.1 Overview of tomographic weak lensing theory

The complex weak lensing shear $\gamma = \gamma_1 + {\rm i}\gamma_2$, which is directly analogous to the complex galaxy ellipticity, can be decomposed into two components: the tangential shear $\gamma_{\rm t}$ and the cross component $\gamma_{\rm x}$. These are defined relative to the separation vector for each pair of galaxies, with $\gamma_{\rm t}$ describing elongation and compression of the ellipticity along the separation vector and $\gamma_{\rm x}$ describing elongation and compression along a direction rotated 45° from the separation vector. The following shear-shear correlation functions

can then be computed:

$$\xi_{\pm}^{\mathrm{k,l}}(\theta) = \frac{\Sigma_{\mathrm{i,j}}\left[\gamma_{\mathrm{t,i}}^{\mathrm{k}}(\vartheta_{\mathrm{i}})\gamma_{\mathrm{t,j}}^{\mathrm{l}}(\vartheta_{\mathrm{j}}) \pm \gamma_{\mathrm{x,i}}^{\mathrm{k}}(\vartheta_{\mathrm{i}})\gamma_{\mathrm{x,j}}^{\mathrm{l}}(\vartheta_{\mathrm{j}})\right] w_{\mathrm{i}} w_{\mathrm{j}} \Delta_{\mathrm{ij}}}{\Sigma_{\mathrm{i,j}} w_{\mathrm{i}} w_{\mathrm{j}} \Delta_{\mathrm{ij}}}, \tag{3.8}$$

where galaxy pairs labelled i,j are separated by angular distance $\vartheta = |\vartheta_{\mathrm{i}} - \vartheta_{\mathrm{j}}|$. If ϑ falls in the angular bin given by θ then Δ_{ij}=1, otherwise Δ_{ij}=0. The labels k,l identify redshift bins. The summation is performed for all galaxies i in bin k and all galaxies j in bin l. The contribution of each galaxy pair is weighted by its inverse variance weight $w_{\mathrm{i}} w_{\mathrm{j}}$. This gives greater significance to galaxy pairs with well-measured shapes.

Shear calibration is performed as described in Miller et al. (2013) and Heymans et al. (2012b). This signal-to-noise (S/N) and size-dependent calibration includes an additive (c) and a multiplicative (m) correction term as follows;

$$\gamma^{\mathrm{obs}} = (1+m)\gamma^{\mathrm{true}} + c. \tag{3.9}$$

An average additive correction of 2×10^{-3} is found for γ_2. The additive correction for γ_1 is found to be consistent with zero. The multiplicative correction to $\xi_{\pm}$ is found by calculating the weighted correlation function of $1+m$ (Miller et al., 2013),

$$1 + \mathrm{K}^{\mathrm{k,l}}(\theta) = \frac{\Sigma_{\mathrm{i,j}}(1+m_{\mathrm{i}}^{\mathrm{k}})(1+m_{\mathrm{j}}^{\mathrm{l}}) w_{\mathrm{i}} w_{\mathrm{j}} \Delta_{\mathrm{ij}}}{\Sigma_{\mathrm{i,j}} w_{\mathrm{i}} w_{\mathrm{j}} \Delta_{\mathrm{ij}}}. \tag{3.10}$$

The shear correlation functions $\xi_{\pm}$ are then corrected by dividing them by $1+\mathrm{K}$.

The shear-shear correlations can also be expressed as filtered functions of the convergence power spectra

$$\xi_{+/-}^{\mathrm{k,l}}(\theta) = \frac{1}{2\pi}\int_0^{\infty} \mathrm{d}\ell\, \ell\, J_{0/4}(\ell\theta) P_{\kappa}^{\mathrm{k,l}}(\ell), \tag{3.11}$$

where J_{n} is the n^{th} order Bessel function of the first kind and ℓ is the modulus of the two-dimensional wave vector. These can be related to line-of-sight integrals of the three-dimensional matter power spectrum

$$P_{\kappa}^{\mathrm{k,l}}(\ell) = \frac{9 H_0^4 \Omega_{\mathrm{m}}^2}{4c^4} \int_0^{\chi_{\mathrm{h}}} \mathrm{d}\chi \frac{g_{\mathrm{k}}(\chi) g_{\mathrm{l}}(\chi)}{a^2(\chi)} P_{\delta}\left(\frac{\ell}{f_K(\chi)}, \chi\right), \tag{3.12}$$

where c is the speed of light, Ω_{m} is the matter energy density, H_0 is the Hubble constant, $f_K(\chi)$ is the comoving angular diameter distance out to a distance χ, χ_{h} is the comoving horizon distance, $a(\chi)$ is the scale factor, and P_{δ} is the 3-dimensional mass power spectrum computed from a non-linear estimation of dark matter clustering (Smith et al., 2003b). The two terms, $g_{\mathrm{k}}(\chi)$, are the geometric lens-efficiency, which depend on the redshift distribution of the sources, $n_{\mathrm{k}}(\chi')$,

$$g_{\mathrm{k}}(\chi) = \int_{\chi}^{\chi_{\mathrm{h}}} \mathrm{d}\chi' n_{\mathrm{k}}(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}. \tag{3.13}$$

Given a cosmological model, matter power spectrum, and redshift distribution of the sources we

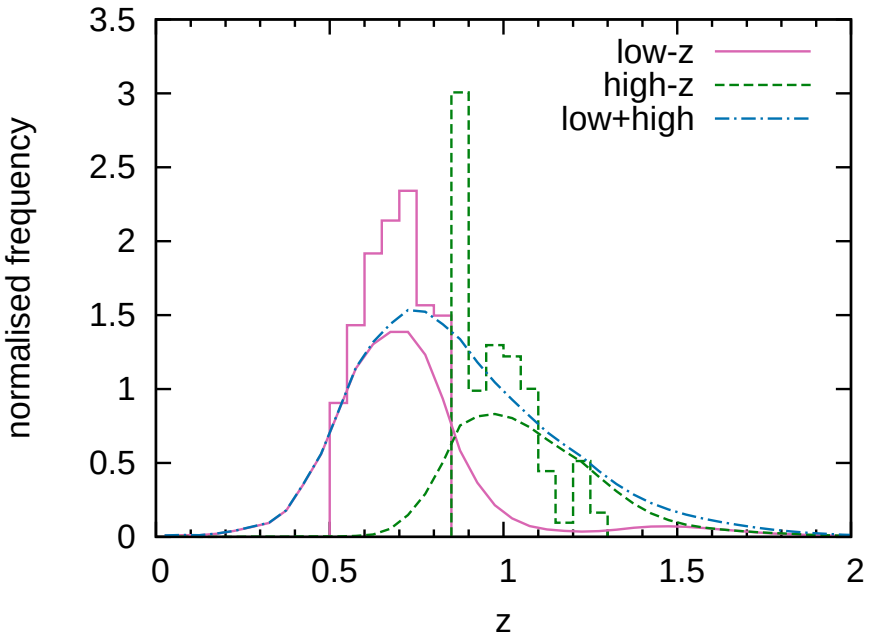

Figure 3.6: Redshift distributions used in the weak lensing analysis. Low and high redshift bins correspond to $z_{\rm p} = (0.5, 0.85]$ and $z_{\rm p} = (0.85, 1.3]$ respectively. Smooth curves show the result of summing the photometric redshift probability distribution functions (PDFs) of all galaxies within the respective redshift bin. The smooth solid and dashed curves are used in the tomographic analysis and the sum of the PDFs over the entire redshift range is given by the smooth dot-dashed line which is used in the 2D lensing analysis. For comparison, the histograms show the redshift distribution obtained from the photometric redshifts.

can model the shear correlation functions. Bayesian model fitting techniques are then used to obtain the posterior probability on the model vector given the observed shear correlation functions. We discuss this further in Section 3.3.3.

3.3.2 The tomographic weak lensing signal

Based on the results presented in Section 3.2, the redshift distribution in each bin is taken to be the sum of the PDFs determined from the photometric redshift analysis of Hildebrandt et al. (2012). We refer to the maximum posterior photometric redshift estimate as the 'photometric redshift'. The histogram of photometric redshifts and the sum of the PDFs for each redshift bin are presented in Figure 3.6. Note that the summed PDFs extend to lower and higher redshifts then the photometric redshifts do, broadening the range below $z = 0.5$ and above $z = 1.3$. The summed PDFs for the two redshift bins also overlap considerably with one another. The average redshift from the summed PDFs is 0.7 for the low redshift bin and 1.05 for the high redshift bin. For the photometric redshifts we find 0.69 and 1.03 for the low and high redshift bins respectively. The average redshift for both bins taken together is found to be 0.87 from the summed PDFs and 0.84 from the photometric redshifts.

We use ATHENA with an opening angle of 0.02 to measure the shear-shear correlation function. We have tested that the difference to the shear-shear correlation function when using an opening angle

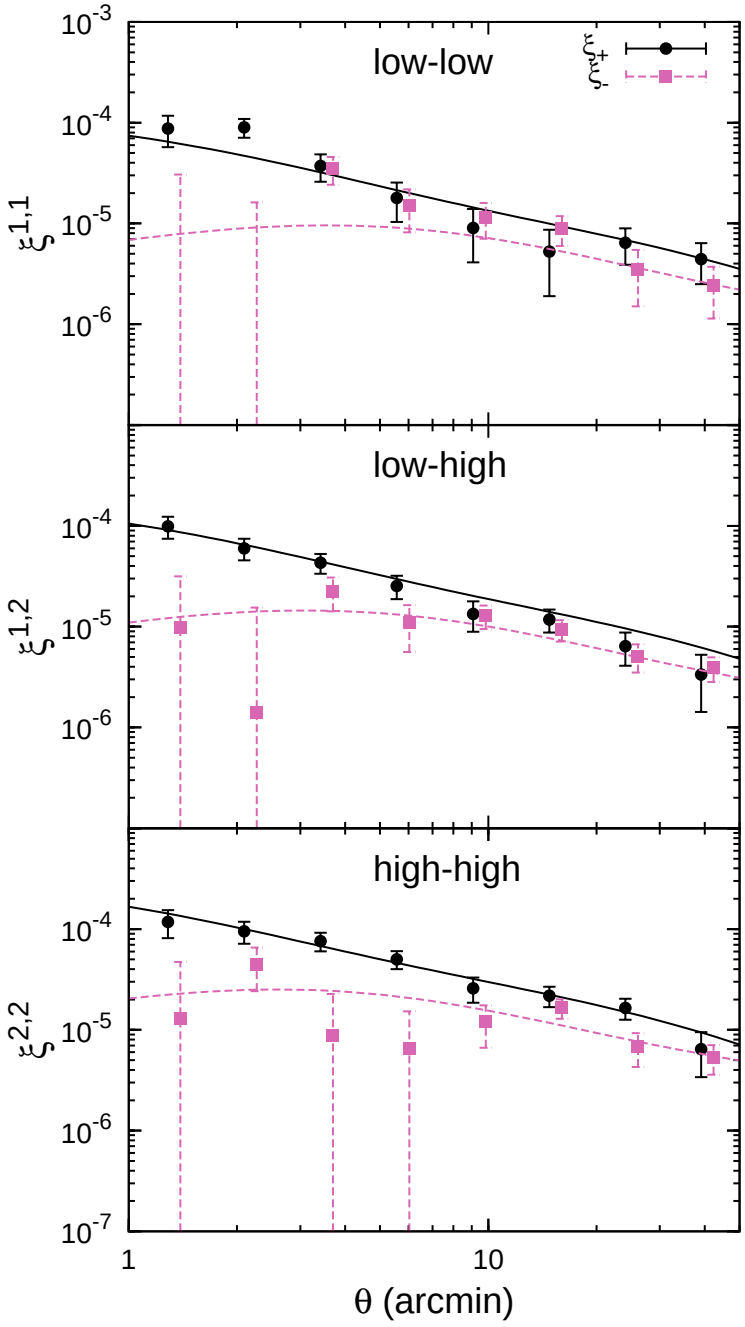

Figure 3.7: The filled circles with solid lines and filled squares with dashed lines show the measured signal for ξ_+ and ξ_- respectively. Each panel shows the shear correlation functions for a unique pairing of redshift bins. The top, middle and bottom panels correspond to low redshift correlated with low redshift (low-low), low with high redshift (low-high), and high with high redshift (high-high). Error bars are the square-root of the diagonal of the covariance matrix measured from mock catalogues (see Section 3.3.3). Theoretical predictions for a fiducial (WMAP7, Komatsu et al., 2011) cosmology are presented as lines; these are not the best-fitting models. There are two negative data points for ξ_- in the top panel, their values are -2.3×10^{-6} and -4.9×10^{-6} for scales 1.34 and 2.18 arcminutes respectively.

Table 3.1: Details of the model dependent cosmological parameters for each of the considered cosmologies. Parameter ranges denote hard priors. A flat distribution is used throughout the range. The bottom three parameters are constrained by WMAP7, and are required in order to deduce σ_8.

Parameter	flat ΛCDM	curved ΛCDM	description
Ω_m	$[0,1.0]$	$[0,1.2]$	Energy density of matter (baryons + dark matter).
σ_8	$[0.2,1.5]$	$[0.2,1.5]$	Normalisation of the matter power spectrum.
h	$[0.4,1.2]$	$[0.4,1.2]$	The dimensionless Hubble constant $h = \frac{H_0}{100\mathrm{km\,s^{-1}\,Mpc^{-1}}}$.
Ω_b	$[0,0.1]$	$[0,0.1]$	Energy density of baryons.
n_s	$[0.7,1.3]$	$[0.7,1.3]$	Slope of the primordial matter power spectrum.
Ω_Λ	$1-\Omega_\mathrm{m}$	$[0,2]$	Energy density of dark energy.
w_0	-1	-1	Constant term in the dark energy equation of state, $w(a) = w_0$.
τ	$[0.04,0.20]$	$[0.04,0.20]$	Reionisation optical depth.
$\Delta^2_\mathscr{R}$	$[1.8,3.5]$	$[1.8,3.5]$	Amplitude of curvature perturbations. Units of 10^{-9} times the amplitude of density fluctuations.
$\mathrm{A_{SZ}}$	$[0.0,2.0]$	$[0.0,2.0]$	Sunyaev-Zel'dovich template amplitude.

of 0.02 compared to a brute force calculation is negligible, approximately 8 per cent of the size of the errors. The signal is first measured on each of the four wide mosaics: W1, W2, W3, and W4, applying the shear calibration described in Section 3.3.1. The correlation functions are then combined by calculating the weighted average. The weight for a given angular bin and wide mosaic is the sum of the inverse variance weight terms for each pair of galaxies. We present ξ_+ and ξ_- for each redshift bin combination in Figure 3.7. The error bars correspond to the diagonal elements from the covariance matrix, discussed in more detail in Section 3.3.3. The lines are the theoretical prediction for a fiducial cosmological model using the Wilkinson Microwave Anisotropy Probe 7-year (WMAP7) best-fitting results (Komatsu et al., 2011), hence the following parameter vector is used: ($\Omega_\mathrm{m} = 0.271$, $\sigma_8 = 0.78$, $h = 0.704$, $\Omega_\mathrm{b} = 0.0455$, $n_\mathrm{s} = 0.967$, $\Omega_\Lambda = 0.729$, $w_0 = -1$). Descriptions of each parameter can be found in Table 3.1. To compute the theoretical models we employ the halo-model of Smith et al. (2003b) to estimate the non-linear matter power spectrum and the analytical approximation of Eisenstein & Hu (1998) to estimate the transfer function.

Emphasizing that no cosmology-dependent systematic tests were used to vet the catalogues (Heymans et al., 2012b), Figure 3.7 demonstrates the robustness of the CFHTLenS catalogues. The tomographic shear signal shows no evidence of a redshift-dependent bias as was seen in earlier CFHTLS data analyses (Kilbinger et al., 2009). We discuss further tests of the redshift scaling of the shear in Section 3.3.3.

3.3.3 Cosmology

From the signal measured in Section 3.3.2, cosmological parameters are estimated using COSMOPMC. COSMOPMC is a freely available[7] Population Monte Carlo (PMC) code, which uses adaptive impor-

[7] http://cosmopmc.info

tance sampling to explore the posterior (Kilbinger et al., 2011). COSMOPMC documentation can be found in Kilbinger et al. (2011); discussion of Bayesian evidence and examples of its application to various cosmological data sets can be found in Kilbinger et al. (2010) and Wraith et al. (2009). The PMC method is detailed in Cappé et al. (2007). The non-linear matter power spectrum is estimated using the halo-model of Smith et al. (2003b). The transfer function is estimated using the analytical approximation of Eisenstein & Hu (1998).

We explore two cosmologies: a flat ΛCDM universe and a curved ΛCDM universe. The model-dependent data vector used with COSMOPMC contains the following seven parameters: $(\Omega_\mathrm{m}, \sigma_8, h, \Omega_\mathrm{b}, n_\mathrm{s}, \Omega_\Lambda, w_0)$. Physical descriptions and priors are presented in Table 3.1. For the flat ΛCDM model we have a five-parameter fit, where we fix $\Omega_\Lambda = 1 - \Omega_\mathrm{m}$ and $w_0 = -1$. For the curved ΛCDM model we have six free parameters as Ω_Λ is allowed to vary, while w_0 remains fixed.

Three other cosmological data sets are used to provide complementary constraining power. Constraints from the cosmic microwave background (CMB) are taken from the 7-year results of WMAP (Komatsu et al., 2011, WMAP7). To obtain parameter constraints we use the publicly released WMAP likelihood code. Baryon acoustic oscillation (BAO) data is taken from the Baryon Oscillation Spectroscopic Survey (Anderson et al., 2012, hereafter referred to as BOSS). We consider the ratio $D_V/r_s = 13.67 \pm 0.22$ of the apparent BAO at $z = 0.57$ to the sound horizon distance to be Gaussian distributed. The Hubble constant is constrained with the results from the HST distance ladder (Riess et al., 2011, hereafter referred to as R11). Following R11, we use a Gaussian prior of mean value $h = 0.738$ and standard deviation $\sigma = 0.024$. For more details of these data sets see Kilbinger et al. (2013). With WMAP7 the parameter set is expanded to include τ, $\mathrm{A_{SZ}}$, and $\Delta^2_{\mathscr{R}}$, from which we deduce σ_8. Prior ranges and brief descriptions are given in Table 3.1. For further details see Komatsu et al. (2011) and references therein. Throughout this section when stating parameter values we quote the 68.3 per cent confidence level as the associated uncertainty with all other parameters marginalised over.

Covariance matrix

In order to estimate a covariance matrix for our measured shear correlation functions in Equation (3.8), we analyse mock CFHTLenS surveys constructed from the three-dimensional N-body numerical lensing simulations of Harnois-Déraps et al. (2012). The 1024^3 particle simulations have a box size of $147.0\,h^{-1}$ Mpc or $231.1\,h^{-1}$ Mpc, depending on the redshift of the simulation, and assume a flat ΛCDM cosmology parametrized by the best-fitting constraints from Komatsu et al. (2009). There are a total of 184 fully independent lines of sight spanning 12.84 square degrees with a resolution of 0.2 arcmin sampled at 26 redshift slices between $0 < z < 3$. The two-point shear statistics match the theoretical predictions of the input cosmology from $0.5 < \theta < 40$ arcmin scales at all redshifts (Harnois-Déraps et al., 2012), this sets the upper angular limit for our tomographic analysis. See Heymans et al. (2012a) for a detailed discussion of covariance matrix estimation from the N-body simulations presented in Harnois-Déraps et al. (2012), including the required Anderson (2003) correction that we apply to debias our estimate of the inverse covariance matrix used in the likelihood analysis that follows.

Table 3.2: Constraints orthogonal to the $\Omega_m - \sigma_8$ degeneracy for a flat ΛCDM cosmology. Results are shown with and without highly non-linear scales which are potentially biased due to non-linear modelling and the effects of baryons (see Section 3.4). 'All scales' refers to scales the correlation functions are measured on: $1 < \theta < 40$ arcmin. We remove scales corresponding to $\xi_- < 10$ arcmin in the case labelled 'removed: $\xi_- < 10$ arcmin'.

Data	$\sigma_8 \left(\frac{\Omega_m}{0.27}\right)^\alpha$	α
tomography:		
all scales	0.771 ± 0.040	0.553 ± 0.016
removed: $\xi_- < 10$ arcmin	0.776 ± 0.041	0.556 ± 0.018
2D Lensing:		
all scales	0.785 ± 0.036	0.556 ± 0.018
removed: $\xi_- < 10$ arcmin	0.780 ± 0.043	0.611 ± 0.015

Flat ΛCDM

We present marginalised two-dimensional likelihood constraints in the $\Omega_m - \sigma_8$ plane in Figure 3.8. The best constraint from weak lensing alone is for a combination of Ω_m and σ_8, which parametrizes the degeneracy. We find $\sigma_8 \left(\frac{\Omega_m}{0.27}\right)^\alpha = 0.771 \pm 0.040$ with $\alpha = 0.553 \pm 0.016$.

When combining CFHTLenS with WMAP7, BOSS, and R11 data sets, we find $\Omega_m = 0.2762 \pm 0.0074$ and $\sigma_8 = 0.802 \pm 0.013$. The precision is ∼20 times better than for CFHTLenS alone where we find $\Omega_m = 0.27 \pm 0.17$ and $\sigma_8 = 0.67 \pm 0.23$. Constraints on the full set of parameters are presented in Table 3.3. We show the results for CFHTLenS tomography, CFHTLenS combined with WMAP7, BOSS and R11, and, to assess the contribution of our data set to these constraints, we include results for WMAP7 combined with R11 and BOSS. The most valuable contribution from CFHTLenS is for Ω_m, σ_8, and Ω_b, where we improve the precision of the constraints by an average factor of 1.5.

For comparison we perform the analysis with a single redshift bin spanning the range of our 2-bin analysis, $0.5 < z_p \leq 1.3$. We refer to this as the 2D lensing case, in contrast to the tomographic case where we split the galaxies into two redshift bins. Figure 3.9 shows the marginalised parameter constraints in the $\Omega_m - \sigma_8$ plane for both 2D lensing and tomography, and the two cases result in very similar constraints. For 2D lensing we find $\sigma_8 \left(\frac{\Omega_m}{0.27}\right)^\alpha = 0.785 \pm 0.036$ and $\alpha = 0.556 \pm 0.018$, which is in agreement with what we find for tomography (Table 3.2). When combining the 2D lensing results from CFHTLenS with WMAP7, BOSS, and R11 data sets, we find $\Omega_m = 0.2774 \pm 0.0074$ and $\sigma_8 = 0.810 \pm 0.013$, which are nearly identical to those found for tomography (listed above and in Table 3.3). This level of agreement is also found for all other parameters when combining CFHTLenS with the other data sets. We note that all the parameter estimates agree within the 68.3 per cent errors and the size of the error bars from 2D lensing are very similar to those found with tomography when combining CFHTLenS with WMAP7, BOSS, and R11 data sets.

For CFHTLenS alone the parameter estimates for 2D lensing and tomographic lensing agree with each other within their 68.3 per cent uncertainties, however, the parameter constraints do not improve.

Table 3.3: Parameter constraints with 68.3 per cent confidence limits. The following parameters are deduced for CFHTLenS: Ω_K and q_0. When combining data sets the deduced parameters are: σ_8, Ω_Λ, and q_0. The label CFHTLenS+Others refers to the combination of CFHTLenS, WMAP7, BOSS, and R11.

Parameter	flat ΛCDM	curved ΛCDM	Data
	0.27 ± 0.17	0.28 ± 0.17	CFHTLenS
Ω_m	0.288 ± 0.010	0.285 ± 0.014	WMAP7+BOSS+R11
	0.2762 ± 0.0074	0.2736 ± 0.0085	CFHTLenS+Others
	0.67 ± 0.23	0.69 ± 0.29	CFHTLenS
σ_8	0.828 ± 0.023	0.819 ± 0.036	WMAP7+BOSS+R11
	0.802 ± 0.013	0.795 ± 0.013	CFHTLenS+Others
	$1 - \Omega_m$	0.38 ± 0.36	CFHTLenS
Ω_Λ	$1 - \Omega_m$	0.717 ± 0.019	WMAP7+BOSS+R11
	$1 - \Omega_m$	0.7312 ± 0.0094	CFHTLenS+Others
	0	0.19 ± 0.43	CFHTLenS
Ω_K	0	-0.0020 ± 0.0061	WMAP7+BOSS+R11
	0	-0.0042 ± 0.0040	CFHTLenS+Others
	0.84 ± 0.25	0.81 ± 0.24	CFHTLenS
h	0.692 ± 0.0088	0.694 ± 0.012	WMAP7+BOSS+R11
	0.6971 ± 0.0081	0.693 ± 0.011	CFHTLenS+Others
	0.030 ± 0.029	0.031 ± 0.030	CFHTLenS
Ω_b	0.0471 ± 0.0012	0.0472 ± 0.0016	WMAP7+BOSS+R11
	0.04595 ± 0.00086	0.0470 ± 0.0015	CFHTLenS+Others
	-0.57 ± 0.27	-0.29 ± 0.40	CFHTLenS
q_0	-0.568 ± 0.016	-0.574 ± 0.025	WMAP7+BOSS+R11
	-0.585 ± 0.011	-0.594 ± 0.014	CFHTLenS+Others
	0.93 ± 0.17	0.91 ± 0.17	CFHTLenS
n_s	0.965 ± 0.012	0.969 ± 0.014	WMAP7+BOSS+R11
	0.960 ± 0.011	0.972 ± 0.012	CFHTLenS+Others
τ	0.086 ± 0.014	0.086 ± 0.015	WMAP7+BOSS+R11
	0.081 ± 0.013	0.085 ± 0.015	CFHTLenS+Others
$\Delta^2_{\mathscr{R}}$	2.465 ± 0.086	2.45 ± 0.13	WMAP7+BOSS+R11
	2.429 ± 0.081	2.361 ± 0.094	CFHTLenS+Others
A_{SZ}	0.97 ± 0.62	1.35 ± 0.61	WMAP7+BOSS+R11
	1.33 ± 0.60	1.39 ± 0.57	CFHTLenS+Others

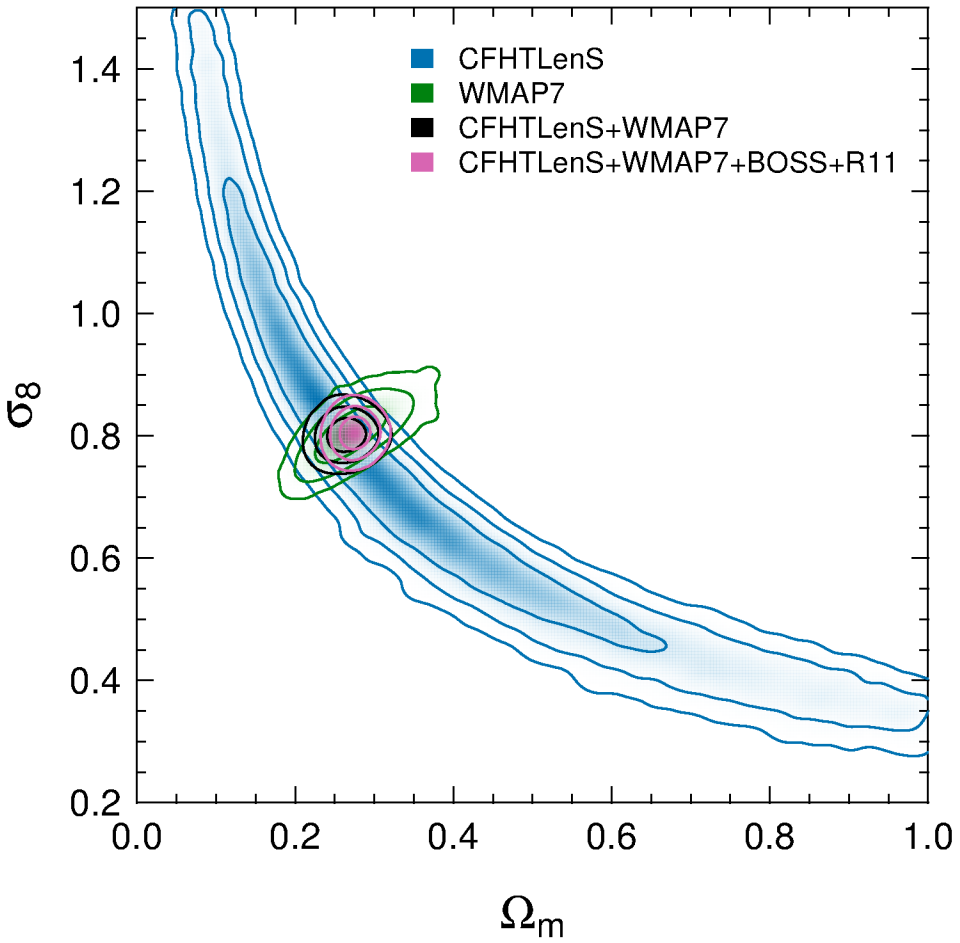

Figure 3.8: Marginalised parameter constraints (68.3, 95.5, and 99.7 per cent confidence levels) in the $\Omega_m - \sigma_8$ plane for a flat ΛCDM model. Results are shown for CFHTLenS (blue), WMAP7 (green), CFHTLenS combined with WMAP7 (black), and CFHTLenS combined with WMAP7, BOSS and R11 (pink).

With two broad overlapping redshift bins of average redshift 0.7 and 1.05, there appears to be insufficient additional information to tighten parameter constraints. Previous estimates of the improvement in constraints from weak lensing tomography use non-overlapping redshift bins, Gaussian covariance, and estimate errors using a Fisher matrix analysis (see Simon et al., 2004). For two redshift bins with $z < 3$ and divided at $z = 0.75$ they find the ratio of the error on individual parameters from tomography to those from a 2D analysis to be $\Delta_{\rm tomo}/\Delta_{\rm 2D} = 0.88$. With our overlapping redshift bins and non-Gaussian covariance it is not surprising that this marginal improvement is significantly degraded. Additionally, our 2D lensing result is for redshifts $0.5 < z_{\rm p} < 1.3$, this removes low redshift galaxies with small signal-to-noise improving constraints and weakening the gains from dividing the redshift range. We therefore expect a modest improvement at best. We find that the ratio of our tomographic errors to our 2D lensing errors is $\Delta_{\rm tomo}/\Delta_{\rm 2D} = 1.16$. The fact that we find larger errors for tomography is surprising and warrants further discussion.

To test our covariance matrices we perform the analysis again, replacing the measured shear correlation function with that predicted from our model using a WMAP7 cosmology. In this case we find that the tomographic errors are a factor of 0.98 of the 2D errors. Therefore the increase in the tomographic errors compared to the 2D errors is not an inherent product of the covariance matrices used. This result confirms that for the case of overlapping redshift bins and non-Gaussian covariance the expected improvement from tomography is marginal, at best.

As discussed in detail in Section 3.4 we have also analysed the data after removing small scales which could be affected by errors in the non-linear modelling of the matter power spectrum and baryonic effects. When removing these scales ($\xi_- < 10$ arcmin) from both the tomographic and 2D lensing analyses we see an improvement in the errors for tomography finding $\Delta_{\rm tomo}/\Delta_{\rm 2D} = 1.04$. The remaining discrepancy could be due to several factors. The Smith et al. (2003b) non-linear prescription could easily be biased at the few per cent level. Residual errors in the redshift of galaxies or other per cent level systematics could be present. In addition we expect some degradation of the tomography errors due to the bias correction of the inverse covariance matrix (Anderson, 2003). The covariance is estimated from a finite number of mock catalogues (see Section 3.3.3), since the tomographic covariance contains three times the number of elements as the 2D covariance, measuring it from the same number of mock catalogues results in a noisier measure. Hartlap et al. (2007) predict an erroneous increase in likelihood area of three per cent given our number of mock catalogues 184 and the size of the data vector for 2D lensing 16 and tomography 48.

Finally we note that our two redshift bins are chosen based on concerns of intrinsic alignment contamination, as such, they are not optimised for constraining cosmology. With more carefully selected redshift bins it may be possible to overcome the issues discussed above and obtain improved cosmological constraints.

Redshift scaling of the cosmic shear signal

Previous CFHTLS data were found to underestimate the shear signal at high redshift necessitating additional calibration parameters when performing cosmological fits to the data (Kilbinger et al., 2009). We demonstrate here that the CFHTLenS data have a redshift dependent shear-signal which agrees with expectations from the modelled ΛCDM cosmology.

The excellent agreement between the 2D and tomographic lensing results (Figure 3.9) suggests that the shear signal across our two redshift bins is scaling as expected. This is also observed in the excellent agreement between the measured shear and the shear prediction based on a fiducial WMAP7 cosmology shown in Figure 3.7.

The shear correlation function for each pair of tomographic redshift bins is analysed separately, corresponding to the shear correlation functions shown in each panel of Figure 3.7. In Figure 3.10 we present marginalised parameter constraints (68.3 per cent confidence level) in the $\Omega_{\rm m} - \sigma_8$ plane for each redshift bin combination. Since each contour is obtained from a sub-sample of the full data-set the degeneracy between the parameters is more pronounced and the area of the contours is larger than when analysing the full data set (Figure 3.8). The agreement between the contours in Figure 3.10 is a

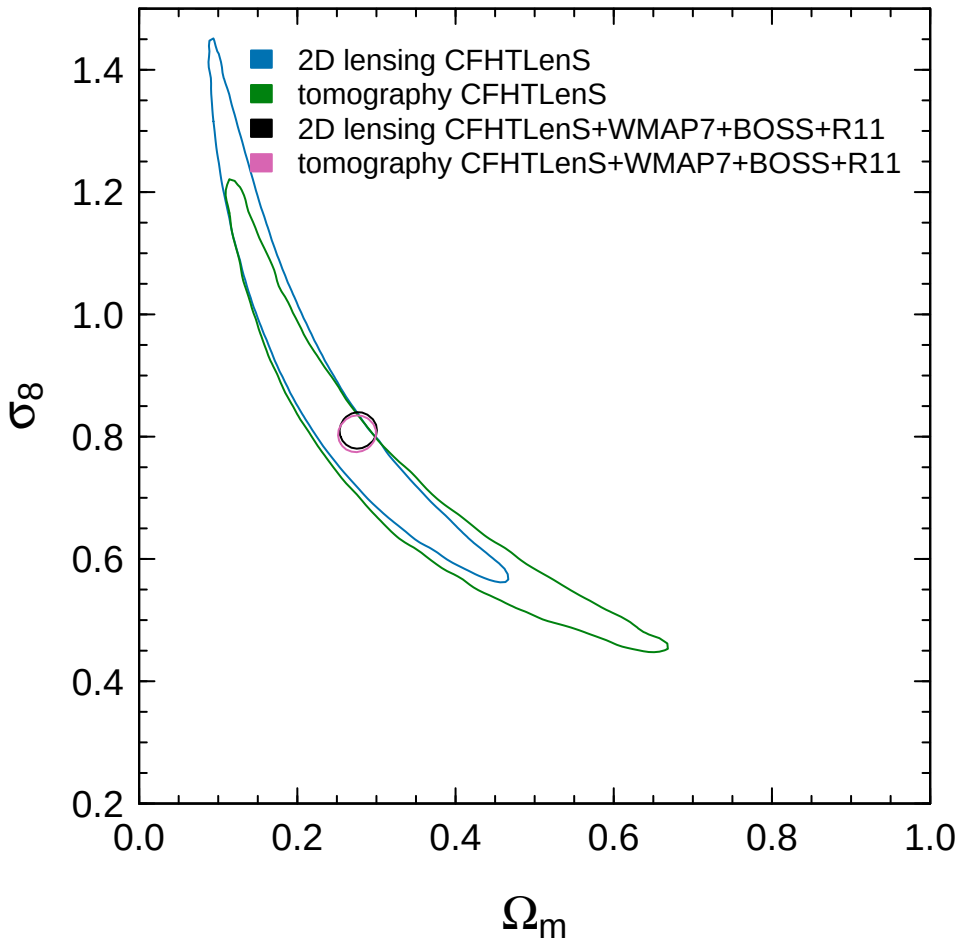

Figure 3.9: Marginalised parameter constraints (68.3 per cent confidence level) in the $\Omega_m - \sigma_8$ plane for a flat ΛCDM cosmology. We compare the results for 2D lensing (blue) and 2-bin tomography (green). We combine CFHTLenS with WMAP7, BOSS, and R11. Results are shown for 2D lensing (black) and 2-bin tomography (pink).

convincing demonstration that the redshift scaling of the shear in the CFHTLenS data is consistent with expectations from the modelled ΛCDM cosmology.

The power-law fits to the degenerate parameter constraints in Figure 3.10 for each case are $\sigma_8 \left(\frac{\Omega_m}{0.27}\right)^{\alpha} = 0.820 \pm 0.067$, 0.753 ± 0.053, and 0.753 ± 0.050 with $\alpha = 0.662 \pm 0.020$, 0.621 ± 0.016, and 0.535 ± 0.013 for the low-low, low-high, and high-high redshift bin pairings respectively.

We reiterate that the cosmological model-dependent verification of redshift scaling presented here is completely independent of the calibration of the data, and the rejection of bad fields, that were done with tests which are not sensitive to cosmology (Heymans et al., 2012b).

Curved ΛCDM

A curved ΛCDM cosmology is modelled, for the full details of parameters and priors used see Table 3.1. We present constraints in the $\Omega_m - \sigma_8$ and $\Omega_m - \Omega_\Lambda$ plane in Figure 3.11. One-dimensional marginalised results when combining CFHTLenS with WMAP7, BOSS, and R11 are $\Omega_m = 0.2736 \pm 0.0085$, $\Omega_\Lambda =$

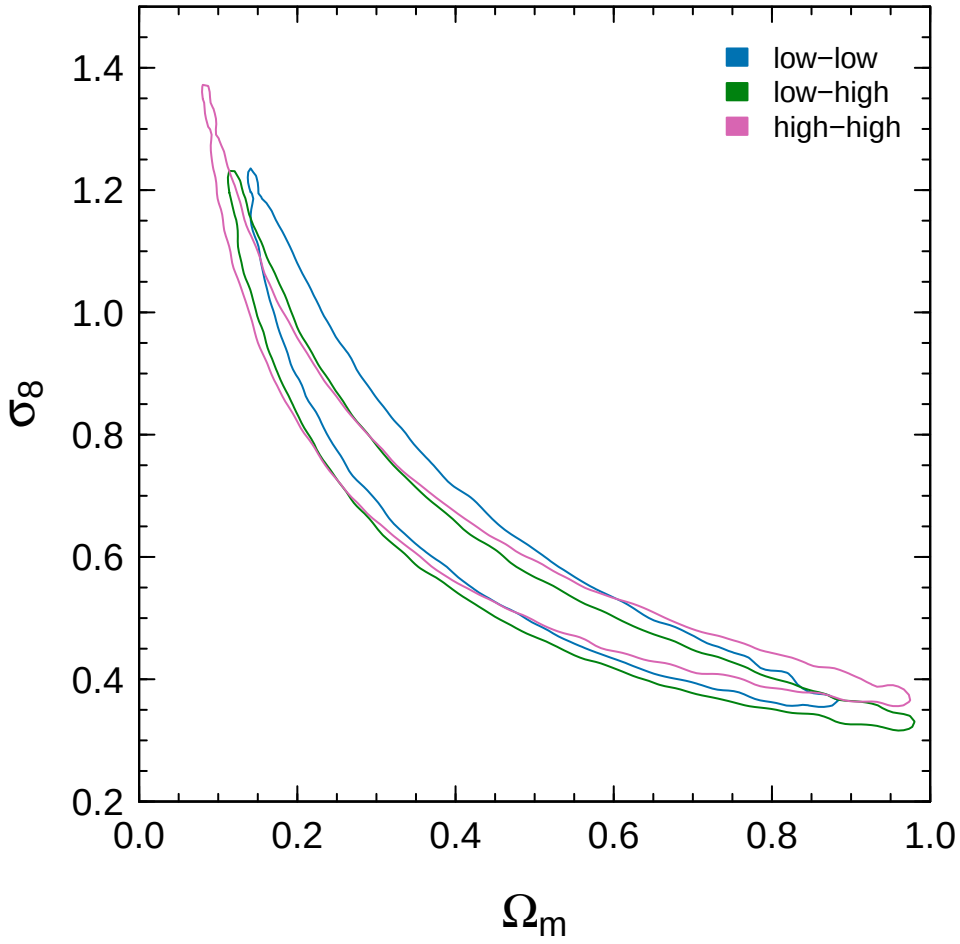

Figure 3.10: Marginalised parameter constraints (68.3 per cent confidence level) in the $\Omega_m - \sigma_8$ plane for a flat ΛCDM cosmology. The results are shown for each combination of the two redshift bins. The low and high redshift bins correspond to $0.5 < z_p \leq 0.85$ and $0.85 < z_p \leq 1.3$ respectively. The excellent agreement shows that redshift scaling of the signal is consistent with the modelled ΛCDM cosmology.

0.7312 ± 0.0094, $\Omega_K = -0.0042 \pm 0.0040$, and $\sigma_8 = 0.795 \pm 0.013$. The constraints on Ω_m and σ_8 do not change significantly from the flat ΛCDM case. Parameter constraints for both models are presented in Table 3.3. The addition of CFHTLenS to WMAP7, BOSS, and R11 is most helpful at constraining Ω_m, σ_8, Ω_K, and Ω_Λ. The precision for these parameters improves, on average, by a factor of two.

We again find excellent agreement with the 2D lensing analysis. When combining the 2D lensing of CFHTLenS with WMAP7, BOSS, and R11 data sets, we find $\Omega_m = 0.2766 \pm 0.0082$, $\Omega_\Lambda = 0.7273 \pm 0.0089$, $\Omega_K = -0.0035 \pm 0.0035$ and $\sigma_8 = 0.804 \pm 0.016$. We do not show the complete details of our 2D lensing parameter estimations. However, we note that in all cases, either with CFHTLenS alone or combined with the other cosmological probes, the 2D results agree with the tomographic results within the 68.3 per cent errors and the size of the error bars are similar for both cases. We again find that for CFHTLenS alone the ratio of individual parameter uncertainties from the tomographic analysis to those of the 2D lensing analysis is $\Delta_{tomo}/\Delta_{2D} = 1.16$.

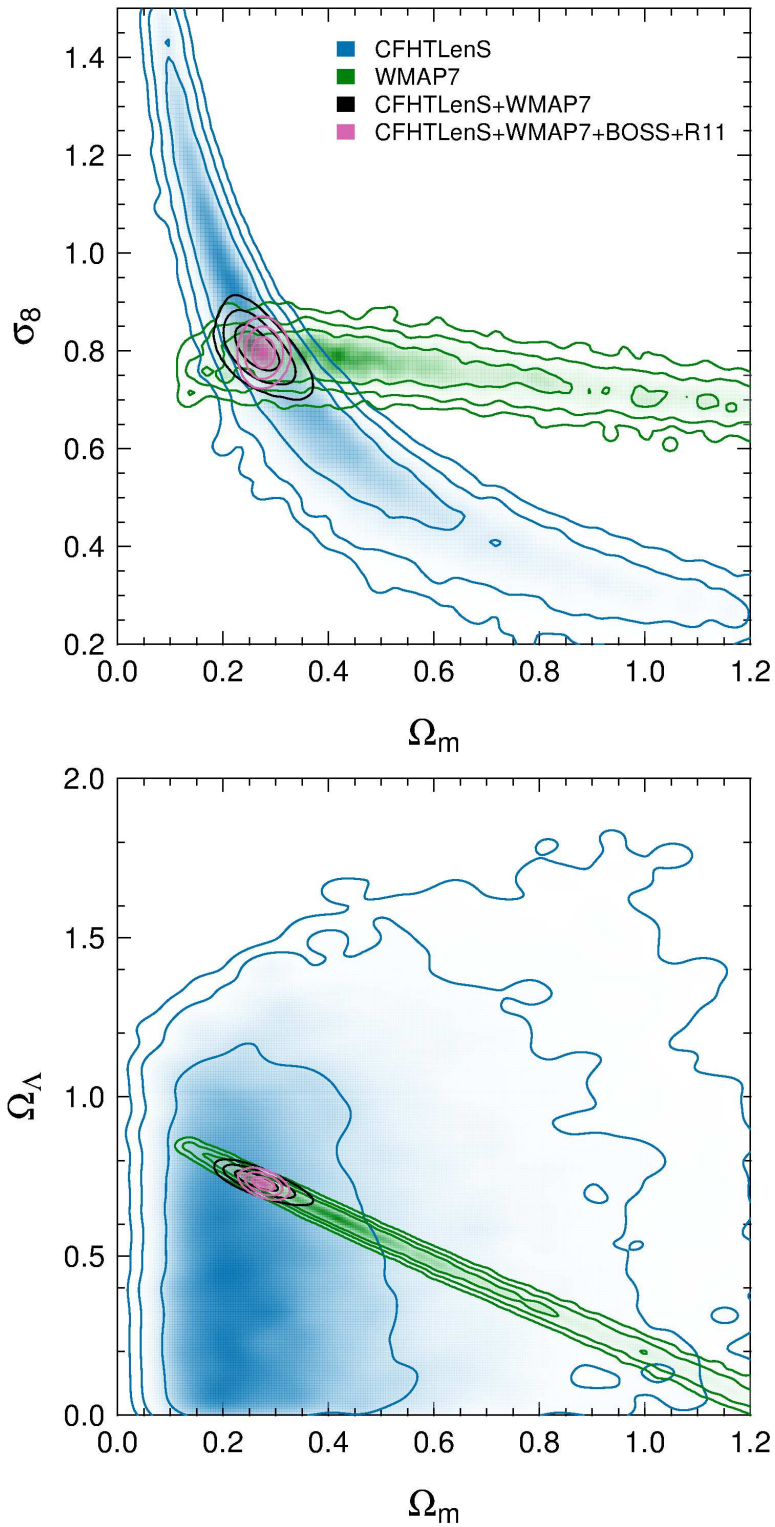

Figure 3.11: Marginalised parameter constraints (68.3, 95.5, and 99.7 per cent confidence levels) for a curved ΛCDM cosmology. Results are shown for CFHTLenS (blue), WMAP7 (green), CFHTLenS combined with WMAP7 (black), and CFHTLenS combined with WMAP7, BOSS, and R11 (pink). **Top panel:** Constraints in the $\Omega_{\rm m} - \sigma_8$ parameter space. **Bottom panel:** Constraints in the $\Omega_{\rm m} - \Omega_\Lambda$ parameter space.

Constraining the deceleration parameter

The deceleration parameter q_0 parametrizes the change in the expansion rate of the Universe. We calculate this as a deduced parameter for both the flat and the curved ΛCDM models. The deceleration parameter depends on the energy density parameters

$$\begin{aligned} q_0 &\equiv -\frac{\ddot{a}(t_0)a(t_0)}{\dot{a}^2(t_0)} = \frac{\Omega_\mathrm{m}}{2} - \Omega_\Lambda \ (\mathrm{curved}\,\Lambda\mathrm{CDM}), \text{ and} \\ &= \frac{3\Omega_\mathrm{m}}{2} - 1 \ (\mathrm{flat}\,\Lambda\mathrm{CDM}), \end{aligned} \tag{3.14}$$

where the scale factor at present time is $a(t_0)$ and derivatives with respect to time are denoted with a dot. For the flat case q_0 is simply a transformation of our results for the matter density parameter Ω_m. We present marginalised constraints for q_0 in Figure 3.12. The pink line is for the curved case where we find $q_0 = -0.29 \pm 0.40$, and the blue line is for the flat case where we find $q_0 = -0.57 \pm 0.27$. Negative values indicate acceleration of the expansion of the Universe. Summing the posterior for $q_0 < 0$ tells us the confidence level at which we have measured an accelerating Universe. For the curved and flat models we find that $q_0 < 0$ at the 82 and 89 per cent confidence level, respectively.

Schrabback et al. (2010) constrain q_0 with a six-bin tomographic analysis of the COSMOS-30 data. Besides having more tomographic bins the redshift range probed is also greater, extending to $z = 4$. For a curved ΛCDM cosmology, holding Ω_b and n_s fixed and using a Gaussian prior on the Hubble constant of $h = 0.72 \pm 0.025$, they find $q_0 < 0$ at 96 per cent confidence. If we do a similar analysis with Ω_b and n_s held fixed and using the a Gaussian prior of $h = 0.738 \pm 0.024$ (R11), we find $q_0 < 0$ at 84 per cent confidence. The difference in constraints on the deceleration parameter can be understood as a result of the much larger values of the dark-energy density preferred by COSMOS-30 $\Omega_\Lambda = 0.97^{+0.39}_{-0.60}$, which lead to smaller values of q_0. Whereas the dark-energy density found here from CFHTLenS is $\Omega_\Lambda = 0.38 \pm 0.36$, resulting in larger values of q_0 for CFHTLenS.

With CFHTLenS alone we are not able to put a strong constraint on the acceleration of the expansion of the Universe. With the addition of the other cosmological probes the entire posterior distribution of q_0 is less than zero. For a curved model with CFHTLenS combined with the other probes, we find $q_0 = -0.594 \pm 0.014$ (see Table 3.3). An accelerating Universe is unambiguously detected.

3.4 Impact of non-linear effects and baryons on the tomographic cosmological constraints

We have presented cosmological parameter constraints from an analysis of the tomographic two-point shear correlation function $\xi_\pm^{\mathrm{k,l}}(\theta)$ (Equation 3.11), incorporating the non-linear dark matter only power spectrum from Smith et al. (2003b) as our theoretical model of $P_\delta(k,z)$ in Equation (3.12). Note that $k = \ell(f_K(\chi))^{-1}$. This halo-model prescription for the non-linear correction has been calibrated on numerical simulations and shown to have an accuracy of $5-10$ per cent over a wide range of scales (Eifler, 2011). The N-body simulations used to estimate the covariance matrices used in this analysis suggest that the accuracy is even better than this for a WMAP5 cosmology (Harnois-Déraps et al., 2012)

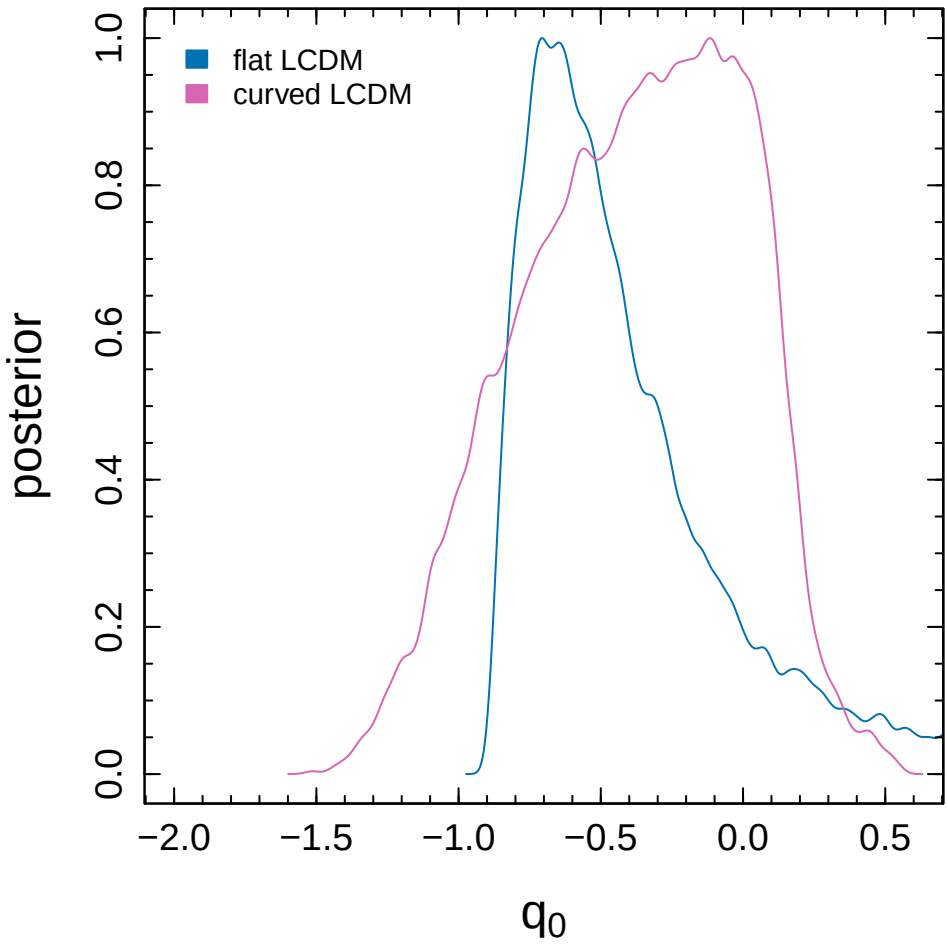

Figure 3.12: Marginalised constraints on the deceleration parameter using the CFHTLenS 2-bin tomographic weak lensing results. An accelerating universe ($q_0 < 0$) is found at the 82 per cent confidence level for a curved ΛCDM model (pink), and at the 89 per cent confidence level for a flat ΛCDM model (blue).

over the redshift range covered in this analysis. While these comparisons give us confidence in our results, and suggest that any error from the non-linear correction will be small in comparison to our statistical error, it is prudent to assess how errors in the non-linear correction will impact our results.

A fully 3D weak lensing analysis of the CFHTLenS data is presented in Kitching et al. (2012). This power spectrum analysis allows for exact redshift dependent cuts in the wave-vector k, which can be motivated by either the comparison of the power spectrum measured from N-body simulations to the non-linear prescription, or the selection of linear scales where the non-linear correction is negligible. For real-space statistics, as used in this chapter and in Kilbinger et al. (2013), it is not possible to make an unambiguous separation of scales. The two-point correlation function $\xi_{\pm}^{\mathrm{k,l}}(\theta)$ is related to the underlying matter power spectrum $P_{\delta}(k,z)$ through integrals over k and z, modulated by the lensing efficiency $g(z)$ (Equation 3.13) and Bessel functions $J_{0/4}(\ell\theta)$ for $\xi_{+/-}(\theta)$ (see Equations 3.11 and 3.12). The measured tomographic two-point shear correlation function $\xi_{\pm}^{\mathrm{k,l}}$ at a fixed scale θ is therefore probing a range of k in the underlying matter power spectrum. In addition, owing to the different Bessel functions, ξ_{+} is

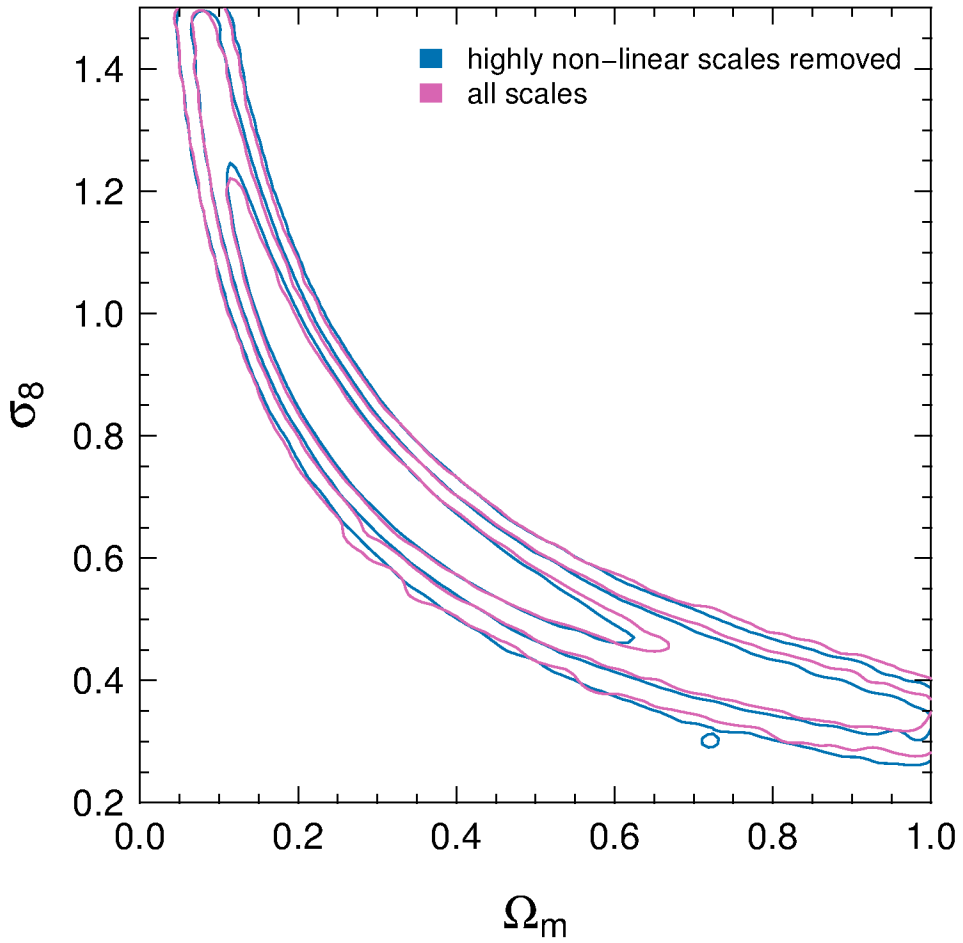

Figure 3.13: Marginalised parameter constraints (68.3, 95. 5, and 99.7 per cent confidence levels) in the $\Omega_{\rm m} - \sigma_8$ plane for a flat ΛCDM cosmology. The pink contours show the result when all eight scales are included, this is the same as the result for CFHTLenS shown in Figure 3.8. The blue contours show the result of removing highly non-linear scales, which are possibly biased due to the non-linear correction to the matter power spectrum or the effect of baryons. We remove the 5 smallest scales of ξ_-, corresponding to $\theta < 10$ arcmin. The contours are only slightly different, indicating that we are not sensitive to these effects given the level of precision of our results.

preferentially probing much smaller k, and hence larger physical scales, than ξ_-.

Kilbinger et al. (2013) present an analysis of the 2D shear correlation function out to large angular scales $\theta < 350$ arcmin. The consistent constraints obtained from the large quasi-linear regime $\theta > 53$ arcmin in comparison to the full angular range, analysed using the Smith et al. (2003b) non-linear power spectrum, give us confidence that the accuracy of this correction is sufficient, falling within our statistical errors.

Comparing the theoretical expectation of $\xi_\pm^{\rm k,l}(\theta)$ (Equation 3.11) for a WMAP7 cosmology, calculated using a non-linear and a linear power spectrum, we determine the angular scale below which the non-linear and linear models differ in amplitude by greater than 10 per cent. For ξ_+, this quasi-linear

limit ranges from $10-14$ arcmin for the three different tomographic combinations (the lowest redshift bin requiring the largest θ cut). For ξ_-, the quasi-linear limit ranges from $100-140$ arcmin. In this analysis, we limit our angular range to scales with $\theta \lesssim 40$ arcmin where we can accurately assess a covariance matrix from the lensing simulations (Harnois-Déraps et al., 2012). We are therefore unable to follow Kilbinger et al. (2013) by limiting our real-space analysis to this quasi-linear regime as we do not probe sufficiently large angular scales. We can however make an assessment of how an error on the non-linear correction would impact our results. We first compare the WMAP7 theoretical expectation of $\xi_\pm^{\mathrm{k,l}}(\theta)$ calculated using a non-linear correction boosted by 7 per cent, with a model calculated with the non-linear correction decreased by 7 per cent. We chose the value of 7 per cent from the average error over the range of k tested in Eifler (2011). We find that these two limits on the non-linear correction cause at least a 10 per cent change in the amplitude of $\xi_\pm^{\mathrm{k,l}}(\theta)$ for scales $\theta \lesssim 1$ arcmin for ξ_+ and $\theta \lesssim 10$ arcmin for ξ_-. Applying these cuts in angular scale corresponds to removing the first 5 angular scales for ξ_- for each tomographic bin shown in Figure 3.7. All ξ_+ scales remain in the analysis. With these scales removed any remaining uncertainty due to the non-linear modelling is well within our statistical error.

For reference, we calculate the approximate wavenumber corresponding to the physical separation of the source galaxies when the small scale cuts are applied. Using the relation $k{=}2\pi(\theta f_K(\chi))^{-1}$, and assuming a flat ΛCDM cosmology, the angular scale cuts correspond to wavenumbers $k_{\theta=1'}{\sim}12\,h\,Mpc^{-1}$ and $k_{\theta=10'}{\sim}1.2\,h\,Mpc^{-1}$. Where we have calculated the angular diameter distance to the mean redshift of the low-redshift bin $f_K(\chi){=}1800\,h^{-1}\,Mpc$. To determine what wavenumber the shear correlation functions are sensitive to we must account for the effect of the Bessel functions, power spectrum, and lensing efficiency in Equations (3.11) and (3.12). To get a rough estimate we take the power spectrum to be a pure power-law with exponent -2, and take the mid-point of the angular diameter distance since this is approximately where the lensing efficiency will peak $f_K(\chi){\sim}900\,h^{-1}\,Mpc$. We perform the integral in Equation (3.11) for both ξ_+ and ξ_- and note for which ℓ 90 per cent of the final value is reached. This results in an approximate wavenumber of $k{\sim}2.5\,h\,Mpc^{-1}$ for both ξ_+ and ξ_-.

Figure 3.13 compares cosmological parameter constraints in the $\Omega_\mathrm{m}-\sigma_8$ plane for this limited number of scales in comparison to the full data set analysed in Section 3.3. The removal of small scales results in a slight change to the degeneracy of the parameters. This test gives us confidence that the non-linear correction used is sufficiently accurate given the statistical error of the survey. This is unlikely to be true for future surveys, where the increased statistical accuracy will require better knowledge of the non-linear correction to the power spectrum (Eifler, 2011).

Finally we turn to the impact of baryons on our results. In our analysis we assume the underlying matter power spectrum is sufficiently well represented by the non-linear dark matter only power spectrum, neglecting the role of baryons. The impact of baryons on the power spectrum is sensitive to the baryonic feedback model used. Therefore, the magnitude of the impact of baryons remains uncertain. Semboloni et al. (2011) present an analysis of cosmological hydrodynamic simulations to quantify the effect of baryon physics on the weak gravitational lensing shear signal using a range of different baryonic feedback models. Their work suggests that a conservative weak lensing analysis should be limited

to those scales where $k \lesssim 1.5\,h\,\mathrm{Mpc}^{-1}$. We implement such a conservative scheme in the 3D power spectrum analysis of Kitching et al. (2012). As discussed above, our real-space analysis mixes k and z scales, leaving us unable to perform a similarly clear test here.

Semboloni et al. (2011) present a comparison of $\xi_{\pm}^{\mathrm{k,l}}(\theta)$ measured for both the cosmological hydrodynamic simulations and a dark matter only simulation for different redshifts, which we use to judge the level of error we should expect baryons to introduce. Assuming the realistic active galactic nucleus feedback model, and considering the scales used in the conservative analysis of Figure 3.13 ($\theta \geq 1.34$ arcmin and $\theta \geq 15.4$ arcmin for ξ_+ and ξ_- respectively), we expect baryons to cause a decrease to the modelled signal of less than ten per cent. The fact that we see very little difference between the conservative and full analysis presented in Figure 3.13 demonstrates that the underlying matter power spectrum is indeed sufficiently well represented by the non-linear dark matter only power spectrum for our statistical accuracy. It also indicates that the impact of baryons on the non-linear dark matter only power spectrum is unlikely to be larger than that predicted by Semboloni et al. (2011). However, baryonic effects will have to be carefully considered for the next generation of weak lensing surveys that will have significantly smaller statistical errors. Semboloni et al. (2012) show the importance of baryonic effects on three-point shear statistics and propose a modification to the modelling of the non-linear matter power spectrum to account for these effects.

3.5 Conclusion

The most important result of this study is that the sum of the photometric redshift probability distribution functions (PDF) within a redshift bin provides an accurate measure of the true redshift distribution of those galaxies; accounting for the scatter due to catastrophic as well as statistical errors. To demonstrate the accuracy of the PDFs we have compared the summed PDFs with the redshift distribution predicted by spectroscopic redshifts, resampled COSMOS-30 redshifts, and predictions from a redshift contamination analysis using the angular correlation function. We find excellent agreement for the redshift range $z_{\mathrm{p}} < 1.3$. This result indicates that the priors and spectral templates used in Hildebrandt et al. (2012) to derive the photometric redshifts provide an accurate and complete description of the galaxies at $z_{\mathrm{p}} < 1.3$. This also motivates our use of the summed PDF as a measure of the redshift distributions in our tomographic weak lensing analysis. Furthermore, the proven accuracy of the summed PDFs provides a reliable method for estimating the source redshift distribution in future weak lensing studies.

We have performed a cosmological analysis of the CFHTLenS data on angular scales $1 < \theta < 40$ arcmin, using two broad redshift bins, $0.5 < z_{\mathrm{p}} \leq 0.85$ and $0.85 < z_{\mathrm{p}} \leq 1.3$, that are not significantly affected by the intrinsic alignment of galaxy shapes. We model two cosmologies; flat and curved ΛCDM. Due to complementary degeneracies our results add valuable constraining power when combined with those from the cosmic microwave background (Komatsu et al., 2011, WMAP7), baryon acoustic oscillations (Anderson et al., 2012, BOSS), and a prior on the Hubble constant (Riess et al., 2011, R11). The addition of our weak lensing results to these other cosmological probes increases the precision of individual marginalised parameter constraints by an average factor of $1.5-2$.

For a flat ΛCDM model the joint parameter constraints for CFHTLenS, WMAP7, BOSS, and R11

are $\Omega_m = 0.2762 \pm 0.0074$ and $\sigma_8 = 0.802 \pm 0.013$. For a curved ΛCDM model, combining the same data sets, we find $\Omega_m = 0.2736 \pm 0.0085$, $\Omega_\Lambda = 0.7312 \pm 0.0094$, $\Omega_K = -0.0042 \pm 0.0040$, and $\sigma_8 = 0.795 \pm 0.013$. Full details of our parameter estimates for both cosmologies are presented in Table 3.3. Our results are consistent with those presented in other studies of the CFHTLenS data: a 2D lensing analysis probing much larger scales where linear theory provides a more accurate model to the matter power spectrum (Kilbinger et al., 2013); and a fine-binned tomographic analysis with six redshift bins accounting for intrinsic alignments (Heymans et al., 2012a).

We compare the tomographic constraints with those from a 2D lensing analysis spanning the same range of redshift $0.5 < z_p \leq 1.3$. We find the two analyses to be completely consistent with all parameter estimates agreeing within their 68.3 per cent confidence levels. We note that the ratio of uncertainties on individual parameters from tomography to those from 2D lensing is on average $\Delta_{\rm tomo}/\Delta_{\rm 2D} = 1.16$. This statistic is 0.98 if we replace our data vectors with a fiducial model, indicating that our covariance matrices do show a slight improvement for tomography. We argue that our non-Gaussian covariance and broad overlapping redshift bins degrade the modest improvement ($\Delta_{\rm tomo}/\Delta_{\rm 2D} = 0.88$) expected from idealised Fisher matrix calculations (Simon et al., 2004). We identify small scales as being largely responsible for the observed increase, finding $\Delta_{\rm tomo}/\Delta_{\rm 2D} = 1.04$ when these scales are removed from the analyses. These scales could be biased due to uncertainties in the modelling of the non-linearities in the matter power spectrum and baryonic effects. Although small scales have inflated our uncertainties from tomography we show in Section 3.4 that they do not significantly affect our results.

Previous analyses of CFHTLS data were hindered by a strong redshift dependent bias in the weak lensing shear, necessitating additional nuisance parameters when analysing the tomographic shear (Kilbinger et al., 2009). We demonstrate that the redshift scaling of the CFHTLenS cosmic shear signal agrees with expectations from the modelled ΛCDM cosmology. The strongest test of this is presented in Figure 3.10, which shows the agreement of cosmological constraints measured for each combination of redshift bins in the $\Omega_m - \sigma_8$ plane for a flat ΛCDM cosmology. This demonstrates the effectiveness of the cosmology-independent tests of residual systematics presented in Heymans et al. (2012b), including the rejection of 25 per cent of the MegaCam pointings which failed to pass these tests. Note also that the shear calibration performed on numerical simulations (Miller et al., 2013) was completed before any cosmological analysis was performed on the data. The two-bin analysis presented here is sensitive to redshift dependent cosmology without introducing additional parameters to model intrinsic alignments, as such it is an excellent final test of the CFHTLenS data product.

Chapter 4

Conclusions and prospects for future research

4.1 Photometric redshift contamination

The analytical framework developed in Chapter 2 is a remarkably useful tool. As outlined in that chapter, there have been several proposals along similar lines (Newman, 2008; Schneider et al., 2006; Zhang et al., 2009). The method developed in this thesis is unique in that only the photometric redshift data are needed. This makes the method applicable to all multi-band optical surveys without the need for extensive spectroscopic overlap (Newman, 2008) or reliance on shear measurements (Zhang et al., 2009). The application of the method in Chapter 3 beautifully demonstrates the utility of this approach. The verification that the sum of the photometric redshifts' Probability Distribution Functions (PDF) provide an accurate measure of the true redshift distribution is extremely useful. Accurate redshift distributions were used in the subsequent weak lensing analysis. This result has been critical in other CFHTLenS analyses that require accurate redshift distributions, including: weak lensing tomography with 6 redshift bins (Heymans et al., 2012a), 2-dimensional weak lensing shear measurements (Kilbinger et al., 2013), tests of general relativity (Simpson et al., 2013), and 2-dimensional mass reconstructions (Van Waerbeke et al., 2013).

Further work can be done to improve the estimation of photometric redshift contamination from the angular correlation function. The most obvious improvement would be to abandon the pair-wise approach wherein the contamination is measured between each pair of bins. A full multi-bin analysis is possible and the necessary theoretical framework is already developed in Chapter 2. The difficulty is the alarming number of free-parameters (N_p) in the analysis as a function of the number of redshift bins (N_z),

$$N_p = N_z(N_z - 1), \tag{4.1}$$

where the number of free-parameters are the contamination fractions $f_{\rm ij}$. For example, the 6-bin analysis presented in Chapter 3 has 30 free-parameters! Simultaneously fitting all of these parameters in finite time becomes a serious challenge. This should be possible with Monte-Carlo Markov Chain (MCMC)

techniques, especially if prior knowledge (e.g. the true number of galaxies within each bin N_i^T) can be used to constrain parameter space. Such an analysis should result in significantly tighter constraints on the parameter space.

In Chapter 3 the sum of the PDFs is used as an estimate of the redshift distribution. No error was assigned to this estimate, and therefore the impact of this uncertainty is missing from the weak lensing error budget. The problem is not in estimating the error, which could easily be done by bootstrapping the summed PDF or by using the standard deviation of the PDF from each galaxy. The largest barrier is in encoding the redshift distribution error in such a way that it is easy to marginalise over. Due to the shape of the summed PDF, simple 3-parameter fitting functions (which have been used in past weak lensing studies) do not fit well. For example, Van Waerbeke et al. (2013) use a 6-parameter model to fit the summed PDF for the redshift bin $0.4 < z < 1.1$. For a tomographic analysis, one would need to add 6 parameters per redshift bin and define ranges over which to allow them to vary. That in itself is problematic, since the derived error distribution will likely not yield a single set of parameter ranges for each of the 6 parameters. An approach like the one used in Benjamin et al. (2007) would be ideal. In that work, a set of 100 redshift distributions, whose parameters were sampled from the full probability distribution, were created and then all 100 redshift distributions were marginalised over. Unfortunately, it is not obvious how to adapt this technique for use with Population Monte-Carlo (PMC) or MCMC codes.

Future work on this method will also need to incorporate the effect of weak lensing magnification, which can cause an angular cross-correlation between redshifts. As discussed in Ménard et al. (2009), dust extinction can also cause an angular cross-correlation on the same order of magnitude as weak lensing magnification; therefore, this effect will also need to be properly accounted for.

4.2 Tomographic weak lensing

The weak lensing tomography analysis presented in Chapter 3 is important for several reasons:

- The scaling of the shear correlation function with redshift is found to agree well with a ΛCDM cosmology. Previous analysis of CFHT Legacy Survey data had to introduce an extra nuisance parameter when fitting the data (Kilbinger et al., 2009). That the redshift scaling works in CFHTLenS is a testament to the thorough systematic testing of the data (Heymans et al., 2012b).

- Removal of the highly non-linear scales is shown to have little impact on the cosmological constraints. Since the interpretation of the weak lensing signal depends on the non-linear modelling of the mass density power spectrum, it is reassuring that the results do not change significantly when they are removed.

- The cosmological constraints demonstrate the power of tomography without the need to model intrinsic alignments (IA). With only two redshift bins, the constraining power is similar to a 6-bin tomographic analysis which is sensitive to potential biases in the modelling of IAs (Heymans et al., 2012a).

- Probing only the quasi-linear regime ($0.1 < \theta < 40$ arcmin), the tomographic analysis is similar in sensitivity to a non-tomographic analysis extending to highly linear scales ($0.9 < \theta < 300$ arcmin, Kilbinger et al., 2013).

- It is shown that a single redshift cut to remove low-z galaxies is extremely powerful, and further sub-dividing the redshift range ($0.5 < z < 1.3$) into two bins slightly decreases parameter constraints. The potency of this '1-bin tomography' was unexpected and future surveys will need to carefully consider how to include these galaxies in their analyses.

One avenue of future work is to determine the impact of low-redshift cuts on cosmological constraints from cosmic shear. A Fisher-information matrix analysis, similar to that presented in Simon et al. (2004), could be done using realistic redshift distributions. The gain in cosmological constraints as a function of the number of redshift bins, and with various low-redshift cuts, would be very informative.

4.2.1 A brief look at Planck

Recent results from the Planck mission (Planck Collaboration XVI et al., 2013) have significantly improved the precision of cosmological constraints from the Cosmic Microwave Background (CMB). Planck surpasses WMAP in its ability to probe small scale anisotropies (i.e. large wavenumbers ℓ). Two complementary data sets that extend to larger ℓ than Planck are the Atacama Cosmology Telescope (ACT, Hasselfield et al., 2013) and the South Pole Telescope (SPT, Reichardt et al., 2013). Planck's constraining power comes from its ability to measure the CMB over the entire sky (but note that galactic masks reduce the total usable area significantly). Although a detailed study of the Planck results cannot be addressed in this thesis, given the importance of the Planck survey a brief discussion is warranted. The current analysis does not include polarization data measured from Planck, yet the size of the error bars are 85 per cent of those from WMAP7, which does include polarization data. The figure of 85 per cent is calculated from the 6 primary parameters of WMAP7, excluding optical depth, which WMAP7 is able to constrain much more tightly due to the inclusion of polarization data.

Figure 4.1 shows the constraints from the Planck-CMB data, $\Omega_{\rm m} = 0.314 \pm 0.02$ and $\sigma_8 = 0.834 \pm 0.027$, over-plotted with the 2-bin tomographic constraints from Figure 3.8. The Planck contours are not derived from their MCMC chains. The contours are unofficial and have been calculated, under the assumption of Gaussianity, from multiples of the 68 per cent confidence limits quoted above. This is a poor substitute for the full likelihood information; however, the degeneracy of Planck's constraints should be very similar to that of WMAP7 (also plotted), suggesting that the extension of the boxes orthogonally to the CFHTLenS constraints is not terribly exaggerated. While the agreement is not perfect, there is not significant tension between the Planck results and those presented in Chapter 3 for CFHTLenS alone. In Figure 4.1 the 95 per cent confidence level of either CFHTLenS or Planck overlaps with the 68 percent confidence level of the other. Note that comparing fully marginalised constraints can be misleading. The tomographic results, $\Omega_{\rm m} = 0.27 \pm 0.17$ and $\sigma_8 = 0.67 \pm 0.23$, suggest much less tension than is evident in Figure 4.1.

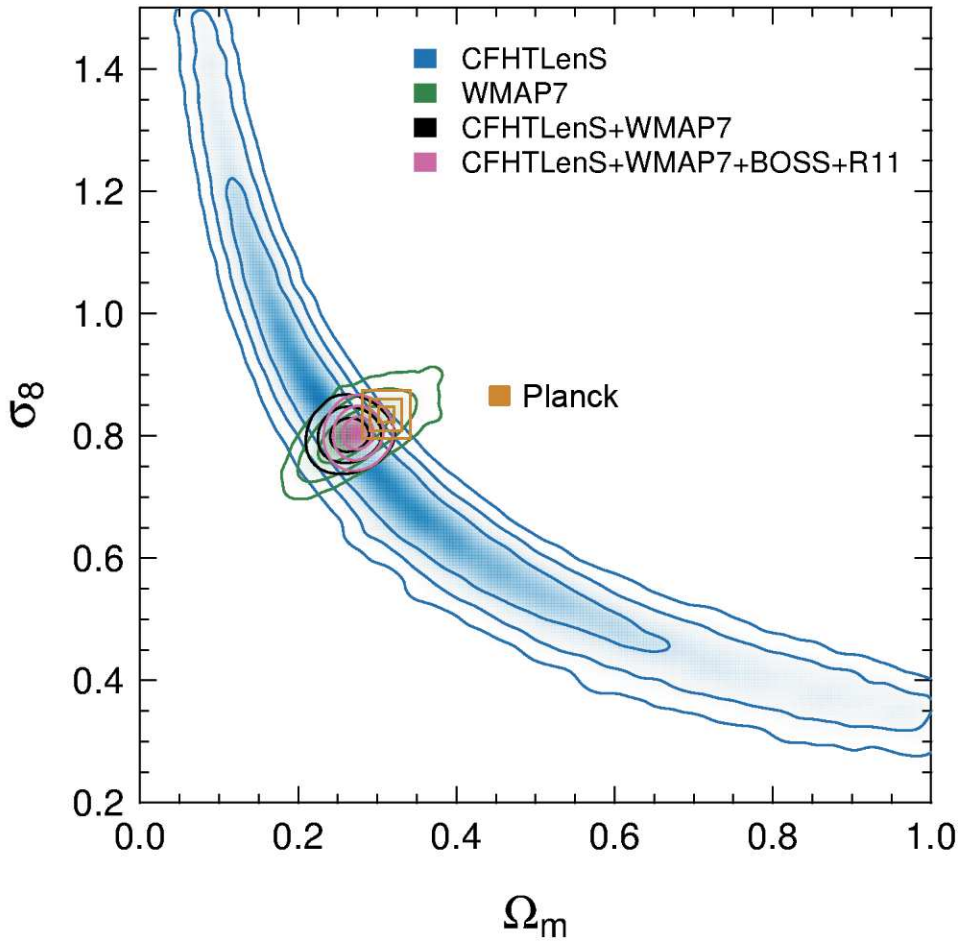

Figure 4.1: Constraints from Figure 3.8 of Chapter 3. The results from Planck are shown to overlap significantly with the 2-bin tomographic results from CFHTLenS. The Planck contours are not official results; they are multiples of the 68 per cent confidence limits given in Planck Collaboration XVI et al. (2013); $\Omega_{\rm m} = 0.314 \pm 0.02$ and $\sigma_8 = 0.834 \pm 0.027$.

Table 4.1 summarizes the parametrized constraint $\sigma_8 \, (\Omega_{\rm m}/0.27)^{\alpha}$ for all three CFHTLenS cosmology papers, the Planck-CMB result discussed above, and the Planck constraint from Sunyaev-Zel'dovich (SZ) cluster counts (Planck-SZ, Planck Collaboration XX et al., 2013). Also included are two other SZ cluster count constraints, one from ACT and the other from SPT. What is striking is the tension between Planck-CMB and all other constraints. In Planck Collaboration XX et al. (2013), it is noted that the SZ cluster constraints can be reconciled if a mass bias parameter of $(1-b) = 0.55 \pm 0.06$ is adopted. This value is typically found to be $(1-b) = 0.8$. It is hard to understand how numerical simulations, X-ray cluster studies, and weak lensing cluster studies could all be similarly biased toward large values of $(1-b)$. No explanation for the discrepancy between Planck-CMB and CFHTLenS results is offered in Planck Collaboration XVI et al. (2013).

Tension between cosmological probes is exciting, and is possibly a hint of physics beyond the standard ΛCDM model. Future work is needed in understanding the tensions that exist between these data

Table 4.1: Constraints on σ_8 for a fixed $\Omega_m = 0.27$, errors are 68 per cent confidence limits. The top three results are from the three CFHTLenS cosmology papers submitted to date. The middle results are from Planck, both the temperature anisotropies of the Cosmic Microwave Background (Planck-CMB) and the Sunyaev-Zel'dovich cluster counts (Planck-SZ). The bottom set of results are from two other Sunyaev-Zel'dovich cluster count measurements made by ACT and SPT. The value of α is not relevant for this comparison but is included for completeness.

Data	$\sigma_8 \left(\frac{\Omega_m}{0.27}\right)^{\alpha}$	α
2-bin tomography (Chapter 3)	0.771 ± 0.040	0.553 ± 0.016
6-bin tomography (Heymans et al., 2012a)	$0.774^{+0.032}_{-0.041}$	0.46 ± 0.02
2D lensing (Kilbinger et al., 2013)	0.787 ± 0.032	0.59 ± 0.02
Planck-CMB (Planck Collaboration XVI et al., 2013)	0.89 ± 0.03	0.46
Planck-SZ (Planck Collaboration XX et al., 2013)	0.782 ± 0.010	0.30
ACT (Hasselfield et al., 2013)	0.768 ± 0.025	0.30
SPT (Reichardt et al., 2013)	0.767 ± 0.037	0.30

sets. If we are fortunate, this tension will not be due to systematic errors in one or several analyses and will instead be a genuine discrepancy due to the different physical processes responsible for these cosmological probes.

4.2.2 Future observations

Future surveys such as the Large Synoptic Survey Telescope (LSST Ivezic et al., 2008), the SuperNova Acceleration Probe (SNAP SNAP Collaboration: G. Aldering et al., 2004) and Euclid (Laureijs et al., 2011) will provide unprecedented precision for weak lensing measurements. LSST is a planned ground based telescope which will survey 30,000 square degrees (20 times the area of CFHTLenS) in 6 photometric bands spanning the optical spectrum. SNAP is a space-based mission optimized to detect Supernovae while conducting a deep wide-field survey. SNAP will survey 10,000 square degrees in 6 photometric bands including three bands in the infrared which will significantly improve the ability to measure photometric redshifts. Finally, Euclid is a planned space-based mission which plans to survey 15,000 square degrees with optical and infrared photometric bands. These surveys will revolutionise the precision of weak lensing measurements. In order to take full advantage of these data, work is needed in understanding non-linear and baryonic effects on the mass density power spectrum. The results in Chapter 3 begin to address these issues.

Bibliography

Albrecht A., Amendola L., Bernstein G., Clowe D., Eisenstein D., Guzzo L., Hirata C., Huterer D., Kirshner R., Kolb E., Nichol R., 2009, ArXiv e-prints: 0901.0721 → pages 48

Albrecht A., Bernstein G., Cahn R., Freedman W. L., Hewitt J., Hu W., Huth J., Kamionkowski M., Kolb E. W., Knox L., Mather J. C., Staggs S., Suntzeff N. B., 2006, ArXiv e-prints: 0609591 → pages 48

Anderson L., Aubourg E., Bailey S., Bizyaev D., Blanton M., Bolton A. S., Brinkmann J., Brownstein J. R., Burden A., Cuesta A. J., da Costa L. A. N., Dawson K. S., de Putter R., Eisenstein D. J., Gunn J. E., Guo H., Hamilton J.-C., Harding P., Ho S., Honscheid K., Kazin E., Kirkby D., Kneib J.-P., Labatie A., Loomis C., Lupton R. H., Malanushenko E., Malanushenko V., Mandelbaum R., Manera M., Maraston C., McBride C. K., Mehta K. T., Mena O., Montesano F., Muna D., Nichol R. C., Nuza S. E., Olmstead M. D., Oravetz D., Padmanabhan N., Palanque-Delabrouille N., Pan K., Parejko J., Pâris I., Percival W. J., Petitjean P., Prada F., Reid B., Roe N. A., Ross A. J., Ross N. P., Samushia L., Sánchez A. G., Schlegel D. J., Schneider D. P., Scóccola C. G., Seo H.-J., Sheldon E. S., Simmons A., Skibba R. A., Strauss M. A., Swanson M. E. C., Thomas D., Tinker J. L., Tojeiro R., Magaña M. V., Verde L., Wagner C., Wake D. A., Weaver B. A., Weinberg D. H., White M., Xu X., Yèche C., Zehavi I., Zhao G.-B., 2012, MNRAS, 427, 3435 (BOSS) → pages ix, 67, 79

Anderson T. W., 2003, An introduction to multivariate statistical analysis. Wiley-Interscience → pages 67, 71

Babul A., Lee M. H., 1991, MNRAS, 250, 407 → pages 7

Bacon D. J., Massey R. J., Refregier A. R., Ellis R. S., 2003, MNRAS, 344, 673 → pages 48

Bacon D. J., Refregier A. R., Ellis R. S., 2000, MNRAS, 318, 625 → pages 7

Bacon D. J., Taylor A. N., Brown M. L., Gray M. E., Wolf C., Meisenheimer K., Dye S., Wisotzki L., Borch A., Kleinheinrich M., 2005, MNRAS, 363, 723 → pages 48

Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291 → pages 7, 48

Benítez N., 2000, ApJ, 536, 571 → pages 18, 49

Benítez N., Ford H., Bouwens R., Menanteau F., Blakeslee J., Gronwall C., Illingworth G., Meurer G., Broadhurst T. J., Clampin M., Franx M., Hartig G. F., Magee D., Sirianni M., Ardila D. R., Bartko F., Brown R. A., Burrows C. J., Cheng E. S., Cross N. J. G., Feldman P. D., Golimowski D. A., Infante L., Kimble R. A., Krist J. E., Lesser M. P., Levay Z., Martel A. R., Miley G. K., Postman M., Rosati P., Sparks W. B., Tran H. D., Tsvetanov Z. I., White R. L., Zheng W., 2004, ApJS, 150, 1 → pages 17

Benjamin J., Heymans C., Semboloni E., Van Waerbeke L., Hoekstra H., Erben T., Gladders M. D., Hetterscheidt M., Mellier Y., Yee H. K. C., 2007, MNRAS, 381, 702 → pages 16, 48, 82

Bernardeau F., Van Waerbeke L., Mellier Y., 1997, A&A, 322, 1 → pages 7

Bernstein G., Huterer D., 2010, MNRAS, 401, 1399 → pages 49

Blandford R. D., Saust A. B., Brainerd T. G., Villumsen J. V., 1991, MNRAS, 251, 600 → pages 7

Bridle S., King L., 2007a, New Journal of Physics, 9, 444 → pages 15

—, 2007b, New Journal of Physics, 9, 444 → pages 62

Brodwin M., Lilly S. J., Porciani C., McCracken H. J., Le Fèvre O., Foucaud S., Crampton D., Mellier Y., 2006, ApJS, 162, 20 → pages 50

Brown M. L., Taylor A. N., Hambly N. C., Dye S., 2002a, MNRAS, 333, 501 → pages 13

—, 2002b, MNRAS, 333, 501 → pages 62

Capak P., Cowie L. L., Hu E. M., Barger A. J., Dickinson M., Fernandez E., Giavalisco M., Komiyama Y., Kretchmer C., McNally C., Miyazaki S., Okamura S., Stern D., 2004, AJ, 127, 180 → pages 49

Cappé O., Douc R., Guillin A., Marin J. M., Robert C. P., 2007, ArXiv e-prints: 0710.4242 → pages 67

Catelan P., Kamionkowski M., Blandford R. D., 2001, MNRAS, 320, L7 → pages 13

Chwolson O., 1924, Astronomische Nachrichten, 221, 329 → pages 5

Crittenden R. G., Natarajan P., Pen U.-L., Theuns T., 2001, ApJ, 559, 552 → pages 13

Croft R. A. C., Metzler C. A., 2000, ApJ, 545, 561 → pages 13, 50

Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, MNRAS, 396, 2379 → pages 50

Dodelson S., 2003, Modern cosmology. Academic Press. → pages 1, 2

Eddington A. S., 1920, Space, time and gravitation. an outline of the general relativity theory. Cambridge: University Press → pages 5

Eifler T., 2011, MNRAS, 418, 536 → pages 75, 78

Einstein A., 1936, Science, 84, 506 → pages 5

Eisenstein D. J., Hu W., 1998, ApJ, 496, 605 → pages 66, 67

Erben T., Hildebrandt H., Lerchster M., Hudelot P., Benjamin J., Van Waerbeke L., Schrabback T., Brimioulle F., Cordes O., Dietrich J. P., Holhjem K., Schirmer M., Schneider P., 2009, A&A, 493, 1197 → pages ix, 21, 23

Erben T., Hildebrandt H., Miller L., Van Waerbeke L., Heymans C., Hoekstra H., Kitching T. D., Mellier Y., Benjamin J., Blake C., Bonnett C., Cordes O., Coupon J., Fu L., Gavazzi R., Gillis B., Grocutt E., Gwyn S. D. J., Holhjem K., Hudson M. J., Kilbinger M., Kuijken K., Milkeraitis M., Rowe B. T. P., Schrabback T., Semboloni E., Simon P., Smit M., Toader O., Vafaei S., van Uitert E., Velander M., 2012, ArXiv e-prints: 1210.8156 → pages 50, 51, 53

Fahlman G., Kaiser N., Squires G., Woods D., 1994, ApJ, 437, 56 → pages 6

Feldmann R., Carollo C. M., Porciani C., Lilly S. J., Capak P., Taniguchi Y., Le Fèvre O., Renzini A., Scoville N., Ajiki M., Aussel H., Contini T., McCracken H., Mobasher B., Murayama T., Sanders D., Sasaki S., Scarlata C., Scodeggio M., Shioya Y., Silverman J., Takahashi M., Thompson D., Zamorani G., 2006, MNRAS, 372, 565 → pages 50

Fort B., Mellier Y., 1994, A&AR, 5, 239 → pages 6

Fort B., Prieur J. L., Mathez G., Mellier Y., Soucail G., 1988, A&A, 200, L17 → pages 6

Fu L., Semboloni E., Hoekstra H., Kilbinger M., Van Waerbeke L., Tereno I., Mellier Y., Heymans C., Coupon J., Benabed K., Benjamin J., Bertin E., Doré O., Hudson M. J., Ilbert O., Maoli R., Marmo C., McCracken H. J., Ménard B., 2008, A&A, 479, 9 → pages 48

Gunn J. E., 1967, ApJ, 150, 737 → pages 7

Hamana T., Miyazaki S., Shimasaku K., Furusawa H., Doi M., Hamabe M., Imi K., Kimura M., Komiyama Y., Nakata F., Okada N., Okamura S., Ouchi M., Sekiguchi M., Yagi M., Yasuda N., 2003, ApJ, 597, 98 → pages 48

Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, ApJ, 374, L1 → pages 10

Harnois-Déraps J., Vafaei S., Van Waerbeke L., 2012, MNRAS, 426, 1262 → pages 67, 75, 78

Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399 → pages 33, 38, 71

Hasselfield M., Hilton M., Marriage T. A., Addison G. E., Barrientos L. F., Battaglia N., Battistelli E. S., Bond J. R., Crichton D., Das S., Devlin M. J., Dicker S. R., Dunkley J., Dunner R., Fowler J. W., Gralla M. B., Hajian A., Halpern M., Hincks A. D., Hlozek R., Hughes J. P., Infante L., Irwin K. D., Kosowsky A., Marsden D., Menanteau F., Moodley K., Niemack M. D., Nolta M. R., Page L. A., Partridge B., Reese E. D., Schmitt B. L., Sehgal N., Sherwin B. D., Sievers J., Sifón C., Spergel D. N., Staggs S. T., Swetz D. S., Switzer E. R., Thornton R., Trac H., Wollack E. J., 2013, ArXiv e-prints: 1301.0816 → pages 83, 85

Hearin A. P., Zentner A. R., Ma Z., Huterer D., 2010, ApJ, 720, 1351 → pages 49

Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649 → pages 13, 50, 62

Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, MNRAS, 347, 895 → pages 15

Heymans C., Grocutt E., Heavens A., Kilbinger M., Kitching T., Simpson F., Benjamin J., Erben T., Hildebrandt H., Hoekstra H., Mellier Y., Van Waerbeke L., Coupon J., Fu L., Harnois-D'eraps J., Hudson M. J., Kilbinger M., Kuijken K., Rowe B., Schrabback T., Semboloni E., Vafaei S., Velander M., 2012a, MNRAS, submitted → pages 50, 62, 67, 80, 81, 82, 85

Heymans C., Van Waerbeke L., Miller L., Erben T., Hildebrandt H., Hoekstra H., Kitching T. D., Mellier Y., Simon P., Bonnett C., Coupon J., Fu L., Harnois Déraps J., Hudson M. J., Kilbinger M., Kuijken K., Rowe B., Schrabback T., Semboloni E., van Uitert E., Vafaei S., Velander M., 2012b, MNRAS, 427, 146 → pages 50, 62, 63, 66, 72, 80, 82

Hildebrandt H., Arnouts S., Capak P., Moustakas L. A., Wolf C., Abdalla F. B., Assef R. J., Banerji M., Benítez N., Brammer G. B., Budavári T., Carliles S., Coe D., Dahlen T., Feldmann R., Gerdes D., Gillis B., Ilbert O., Kotulla R., Lahav O., Li I. H., Miralles J. M., Purger N., Schmidt S., Singal J., 2010, A&A, 523, A31 → pages 17

Hildebrandt H., Erben T., Kuijken K., Van Waerbeke L., Heymans C., Coupon J., Benjamin J., Bonnett C., Fu L., Hoekstra H., Kitching T. D., Mellier Y., Miller L., Velander M., Hudson M. J., Rowe B. T. P., Schrabback T., Semboloni E., Benítez N., 2012, MNRAS, 421, 2355 → pages 18, 19, 49, 50, 51, 52, 54, 55, 56, 61, 64, 79

Hildebrandt H., Pielorz J., Erben T., Van Waerbeke L., Simon P., Capak P., 2009, A&A, 498, 725 → pages 55

Hirata C. M., Mandelbaum R., Ishak M., Seljak U., Nichol R., Pimbblet K. A., Ross N. P., Wake D., 2007, MNRAS, 381, 1197 → pages 15

Hirata C. M., Seljak U., 2004a, PhRvD, 70, 063526 → pages 15

—, 2004b, PhRvD, 70, 063526 → pages 50, 62

Hoekstra H., Mellier Y., Van Waerbeke L., Semboloni E., Fu L., Hudson M. J., Parker L. C., Tereno I., Benabed K., 2006, ApJ, 647, 116 → pages 48

Hu W., 1999, ApJ, 522, L21 → pages 48

Hu W., Tegmark M., 1999, ApJ, 514, L65 → pages 7

Huterer D., 2002, PhRvD, 65, 063001 → pages 48

Huterer D., Kim A., Krauss L. M., Broderick T., 2004, ApJ, 615, 595 → pages 21

Ilbert O., Arnouts S., McCracken H. J., Bolzonella M., Bertin E., Le Fèvre O., Mellier Y., Zamorani G., Pellò R., Iovino A., Tresse L., Le Brun V., Bottini D., Garilli B., Maccagni D., Picat J. P., Scaramella R., Scodeggio M., Vettolani G., Zanichelli A., Adami C., Bardelli S., Cappi A., Charlot S., Ciliegi P., Contini T., Cucciati O., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Marano B., Marinoni C., Mazure A., Meneux B., Merighi R., Paltani S., Pollo A., Pozzetti L., Radovich M., Zucca E., Bondi M., Bongiorno A., Busarello G., de La Torre S., Gregorini L., Lamareille F., Mathez G., Merluzzi P., Ripepi V., Rizzo D., Vergani D., 2006, A&A, 457, 841 → pages 38, 40, 46, 52

Ilbert O., Capak P., Salvato M., Aussel H., McCracken H. J., Sanders D. B., Scoville N., Kartaltepe J., Arnouts S., Le Floc'h E., Mobasher B., Taniguchi Y., Lamareille F., Leauthaud A., Sasaki S., Thompson D., Zamojski M., Zamorani G., Bardelli S., Bolzonella M., Bongiorno A., Brusa M., Caputi K. I., Carollo C. M., Contini T., Cook R., Coppa G., Cucciati O., de la Torre S., de Ravel L., Franzetti P., Garilli B., Hasinger G., Iovino A., Kampczyk P., Kneib J. P., Knobel C., Kovac K., Le Borgne J. F., Le Brun V., Fèvre O. L., Lilly S., Looper D., Maier C., Mainieri V., Mellier Y., Mignoli M., Murayama T., Pellò R., Peng Y., Pérez-Montero E., Renzini A., Ricciardelli E., Schiminovich D., Scodeggio M., Shioya Y., Silverman J., Surace J., Tanaka M., Tasca L., Tresse L., Vergani D., Zucca E., 2009, ApJ, 690, 1236 → pages 49, 55

Ivezic Z., Tyson J. A., Acosta E., Allsman R., Anderson S. F., Andrew J., Angel R., Axelrod T., Barr J. D., Becker A. C., Becla J., Beldica C., Blandford R. D., Bloom J. S., Borne K., Brandt W. N., Brown M. E., Bullock J. S., Burke D. L., Chandrasekharan S., Chesley S., Claver C. F., Connolly A., Cook K. H., Cooray A., Covey K. R., Cribbs C., Cutri R., Daues G., Delgado F., Ferguson H., Gawiser E., Geary J. C., Gee P., Geha M., Gibson R. R., Gilmore D. K., Gressler W. J., Hogan C., Huffer M. E., Jacoby S. H., Jain B., Jernigan J. G., Jones R. L., Juric M., Kahn S. M., Kalirai J. S., Kantor J. P., Kessler R., Kirkby D., Knox L., Krabbendam V. L., Krughoff S., Kulkarni S., Lambert R., Levine D., Liang M., Lim K., Lupton R. H., Marshall P., Marshall S., May M., Miller M., Mills D. J., Monet D. G., Neill D. R., Nordby M., O'Connor P., Oliver J., Olivier S. S., Olsen K., Owen R. E., Peterson J. R., Petry C. E., Pierfederici F., Pietrowicz S., Pike R., Pinto P. A., Plante R., Radeka V., Rasmussen A., Ridgway S. T., Rosing W., Saha A., Schalk T. L., Schindler R. H., Schneider D. P., Schumacher G., Sebag J., Seppala L. G., Shipsey I., Silvestri N., Smith J. A., Smith R. C., Strauss M. A., Stubbs C. W., Sweeney D., Szalay A., Thaler J. J., Vanden Berk D., Walkowicz L., Warner M., Willman B., Wittman D., Wolff S. C., Wood-Vasey W. M., Yoachim P., Zhan H., for the LSST Collaboration, 2008, ArXiv e-prints: 0805.2366 → pages 85

Jain B., Seljak U., 1997, ApJ, 484, 560 → pages 7

Jaroszynski M., Park C., Paczynski B., Gott III J. R., 1990, ApJ, 365, 22 → pages 7

Jarvis M., Bernstein G. M., Fischer P., Smith D., Jain B., Tyson J. A., Wittman D., 2003, AJ, 125, 1014 → pages 48

Joachimi B., Bridle S. L., 2010, A&A, 523, A1 → pages 15

Joachimi B., Mandelbaum R., Abdalla F. B., Bridle S. L., 2011, A&A, 527, A26 → pages 15

Kaiser N., 1992, ApJ, 388, 272 → pages 7, 10

—, 1998, ApJ, 498, 26 → pages 7

Kaiser N., Squires G., 1993, ApJ, 404, 441 → pages 6

Kaiser N., Wilson G., Luppino G. A., 2000, ArXiv Astrophysics e-prints → pages 7

Kamionkowski M., Babul A., Cress C. M., Refregier A., 1998, MNRAS, 301, 1064 → pages 7

Kerscher M., Szapudi I., Szalay A. S., 2000, ApJ, 535, L13 → pages 22

Kilbinger M., Benabed K., Cappe O., Cardoso J. F., Coupon J., Fort G., McCracken H. J., Prunet S., Robert C. P., Wraith D., 2011, ArXiv e-prints: 1101.0950 → pages 67

Kilbinger M., Benabed K., Guy J., Astier P., Tereno I., Fu L., Wraith D., Coupon J., Mellier Y., Balland C., Bouchet F. R., Hamana T., Hardin D., McCracken H. J., Pain R., Regnault N., Schultheis M., Yahagi H., 2009, A&A, 497, 677 → pages 50, 62, 66, 71, 80, 82

Kilbinger M., Fu L., Heymans C., Simpson F., Benjamin J., Erben T., Harnois-Déraps J., Hoekstra H., Hildebrandt H., Kitching T. D., Mellier Y., Miller L., Van Waerbeke L., Benabed K., Bonnett C., Coupon J., Hudson M. J., Kuijken K., Rowe B., Schrabback T., Semboloni E., Vafaei S., Velander M., 2013, MNRAS, 735 → pages 50, 67, 76, 77, 78, 80, 81, 83, 85

Kilbinger M., Wraith D., Robert C. P., Benabed K., Cappé O., Cardoso J. F., Fort G., Prunet S., Bouchet F. R., 2010, MNRAS, 405, 2381 → pages 67

King L. J., 2005, A&A, 441, 47 → pages 15

King L. J., Schneider P., 2003, A&A, 398, 23 → pages 15

Kirk D., Bridle S., Schneider M., 2010, MNRAS, 408, 1502 → pages 15

Kitching T. D., Heavens A., CFHTLenS Collaboration, 2012, In preparation → pages 76, 79

Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2 → pages 28, 31

Klimov Y. G., 1963, Soviet Physics Doklady, 8, 119 → pages 5

Komatsu E., Dunkley J., Nolta M. R., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., Limon M., Page L., Spergel D. N., Halpern M., Hill R. S., Kogut A., Meyer S. S., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2009, ApJS, 180, 330 → pages 67

Komatsu E., Smith K. M., Dunkley J., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., Nolta M. R., Page L., Spergel D. N., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Odegard N., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2011, ApJS, 192, 18 (WMAP7) → pages 65, 66, 67, 79

Kristian J., Sachs R. K., 1966, ApJ, 143, 379 → pages 7

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64 → pages 22, 56

Laureijs R., Amiaux J., Arduini S., Augueres J. ., Brinchmann J., Cole R., Cropper M., Dabin C., Duvet L., Ealet A., et al., 2011, ArXiv e-prints: 1110.3193 → pages 85

Le Fèvre O., Vettolani G., Garilli B., Tresse L., Bottini D., Le Brun V., Maccagni D., Picat J. P., Scaramella R., Scodeggio M., Zanichelli A., Adami C., Arnaboldi M., Arnouts S., Bardelli S., Bolzonella M., Cappi A., Charlot S., Ciliegi P., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Ilbert O., Iovino A., McCracken H. J., Marano B., Marinoni C., Mathez G., Mazure A., Meneux B., Merighi R., Paltani S., Pellò R., Pollo A., Pozzetti L., Radovich M., Zamorani G., Zucca E., Bondi M., Bongiorno A., Busarello G., Lamareille F., Mellier Y., Merluzzi P., Ripepi V., Rizzo D., 2005, A&A, 439, 845 → pages 53

Lee J., Pen U.-L., 2001, ApJ, 555, 106 → pages 13

Lee M. H., Paczynski B., 1990, ApJ, 357, 32 → pages 7

Liebes S., 1964, Physical Review, 133, 835 → pages 5

Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118 → pages 50

Lynds R., Petrosian V., 1986, in Bulletin of the American Astronomical Society, Vol. 18, Bulletin of the American Astronomical Society, p. 1014 → pages 6

Ma Z., Hu W., Huterer D., 2006, ApJ, 636, 21 → pages 49

Mandelbaum R., Blake C., Bridle S., Abdalla F. B., Brough S., Colless M., Couch W., Croom S., Davis T., Drinkwater M. J., Forster K., Glazebrook K., Jelliffe B., Jurek R. J., Li I.-H., Madore B., Martin C., Pimbblet K., Poole G. B., Pracy M., Sharp R., Wisnioski E., Woods D., Wyder T., 2011, MNRAS, 410, 844 → pages 15

Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, MNRAS, 367, 611 → pages 15, 16

Mandelbaum R., Seljak U., Hirata C. M., Bardelli S., Bolzonella M., Bongiorno A., Carollo M., Contini T., Cunha C. E., Garilli B., Iovino A., Kampczyk P., Kneib J.-P., Knobel C., Koo D. C., Lamareille F., Le Fèvre O., Le Borgne J.-F., Lilly S. J., Maier C., Mainieri V., Mignoli M., Newman J. A., Oesch P. A., Perez-Montero E., Ricciardelli E., Scodeggio M., Silverman J., Tasca L., 2008, MNRAS, 386, 781 → pages 50

Massey R., Rhodes J., Leauthaud A., Capak P., Ellis R., Koekemoer A., Réfrégier A., Scoville N., Taylor J. E., Albert J., Bergé J., Heymans C., Johnston D., Kneib J. P., Mellier Y., Mobasher B., Semboloni E., Shopbell P., Tasca L., Van Waerbeke L., 2007, ApJS, 172, 239 → pages 48

Mellier Y., Fort B., Soucail G., Mathez G., Cailloux M., 1991, ApJ, 380, 334 → pages 6

Ménard B., Scranton R., Fukugita M., Richards G., 2009, ArXiv:0902.4240 → pages 46, 82

Miller L., Heymans C., Kitching T. D., van Waerbeke L., Erben T., Hildebrandt H., Hoekstra H., Mellier Y., Rowe B. T. P., Coupon J., Dietrich J. P., Fu L., Harnois-Déraps J., Hudson M. J., Kilbinger M., Kuijken K., Schrabback T., Semboloni E., Vafaei S., Velander M., 2013, MNRAS, 429, 2858 → pages 9, 50, 55, 63, 80

Newman J. A., 2008, ArXiv:0805.1409 → pages 21, 81

Newman J. A., Cooper M. C., Davis M., Faber S. M., Coil A. L., Guhathakurta P., Koo D. C., Phillips A. C., Conroy C., Dutton A. A., Finkbeiner D. P., Gerke B. F., Rosario D. J., Weiner B. J., Willmer C. N. A., Yan R., Harker J. J., Kassin S. A., Konidaris N. P., Lai K., Madgwick D. S., Noeske K. G., Wirth G. D., Connolly A. J., Kaiser N., Kirby E. N., Lemaux B. C., Lin L., Lotz J. M., Luppino G. A., Marinoni C., Matthews D. J., Metevier A., Schiavon R. P., 2012, ArXiv e-prints: 1203.3192 → pages 53

Paczynski B., 1987, Nature, 325, 572 → pages 6

Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19 → pages 10, 12, 14

Peacock J. A., Schneider P., Efstathiou G., Ellis J. R., Leibundgut B., Lilly S. J., Mellier Y., 2006, ESA-ESO Working Group on "Fundamental Cosmology". Tech. rep., this is an institution → pages 48

Peebles P. J. E., 1980, The large-scale structure of the universe. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p. → pages 22

—, 1993, Principles of Physical Cosmology. Princeton University Press → pages 11

Pen U.-L., Lee J., Seljak U., 2000, ApJ, 543, L107 → pages 13

Perlmutter S., Aldering G., Goldhaber G., Knop R. A., Nugent P., Castro P. G., Deustua S., Fabbro S., Goobar A., Groom D. E., Hook I. M., Kim A. G., Kim M. Y., Lee J. C., Nunes N. J., Pain R., Pennypacker C. R., Quimby R., Lidman C., Ellis R. S., Irwin M., McMahon R. G., Ruiz-Lapuente P., Walton N., Schaefer B., Boyle B. J., Filippenko A. V., Matheson T., Fruchter A. S., Panagia N., Newberg H. J. M., Couch W. J., Supernova Cosmology Project, 1999, ApJ, 517, 565 → pages 2

Petters A. O., Levine H., Wambsganss J., 2001, Singularity theory and gravitational lensing. Progress in mathematical physics; vol 21 → pages 6

Planck Collaboration XVI, Ade P. A. R., Aghanim N., Armitage-Caplan C., Arnaud M., Ashdown M., Atrio-Barandela F., Aumont J., Baccigalupi C., Banday A. J., et al., 2013, ArXiv e-prints: 1303.5076 → pages 83, 84, 85

Planck Collaboration XX, Ade P. A. R., Aghanim N., Armitage-Caplan C., Arnaud M., Ashdown M., Atrio-Barandela F., Aumont J., Baccigalupi C., Banday A. J., et al., 2013, ArXiv e-prints: 1303.5080 → pages 84, 85

Refregier A., 2003, ARA&A, 41, 645 → pages 7

Refsdal S., 1964a, MNRAS, 128, 307 → pages 5

—, 1964b, MNRAS, 128, 295 → pages 5

Reichardt C. L., Stalder B., Bleem L. E., Montroy T. E., Aird K. A., Andersson K., Armstrong R., Ashby M. L. N., Bautz M., Bayliss M., Bazin G., Benson B. A., Brodwin M., Carlstrom J. E., Chang C. L., Cho H. M., Clocchiatti A., Crawford T. M., Crites A. T., de Haan T., Desai S., Dobbs M. A., Dudley J. P., Foley R. J., Forman W. R., George E. M., Gladders M. D., Gonzalez A. H., Halverson N. W., Harrington N. L., High F. W., Holder G. P., Holzapfel W. L., Hoover S., Hrubes J. D., Jones C., Joy M., Keisler R., Knox L., Lee A. T., Leitch E. M., Liu J., Lueker M., Luong-Van D., Mantz A., Marrone D. P., McDonald M., McMahon J. J., Mehl J., Meyer S. S., Mocanu L., Mohr J. J., Murray S. S., Natoli T., Padin S., Plagge T., Pryke C., Rest A., Ruel J., Ruhl J. E., Saliwanchik B. R., Saro A., Sayre J. T., Schaffer K. K., Shaw L., Shirokoff E., Song J., Spieler H. G., Staniszewski Z., Stark A. A., Story K., Stubbs C. W., Šuhada R., van Engelen A., Vanderlinde K., Vieira J. D., Vikhlinin A., Williamson R., Zahn O., Zenteno A., 2013, ApJ, 763, 127 → pages 83, 85

Riess A. G., Filippenko A. V., Challis P., Clocchiatti A., Diercks A., Garnavich P. M., Gilliland R. L., Hogan C. J., Jha S., Kirshner R. P., Leibundgut B., Phillips M. M., Reiss D., Schmidt B. P., Schommer R. A., Smith R. C., Spyromilio J., Stubbs C., Suntzeff N. B., Tonry J., 1998, AJ, 116, 1009 → pages 2

Riess A. G., Macri L., Casertano S., Lampeitl H., Ferguson H. C., Filippenko A. V., Jha S. W., Li W., Chornock R., 2011, ApJ, 730, 119 (R11) → pages x, xi, 67, 79

Ryden B., 2003, Introduction to cosmology. Addison Wesley → pages 2

Schneider M., Knox L., Zhan H., Connolly A., 2006, ApJ, 651, 14 → pages 21, 81

Schneider P., 2005, ArXiv Astrophysics e-prints → pages 7, 9

Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer-Verlag → pages 6

Schneider P., Weiss A., 1988, ApJ, 327, 526 → pages 7

Schrabback T., Hartlap J., Joachimi B., Kilbinger M., Simon P., Benabed K., Bradač M., Eifler T., Erben T., Fassnacht C. D., High F. W., Hilbert S., Hildebrandt H., Hoekstra H., Kuijken K., Marshall P. J., Mellier Y., Morganson E., Schneider P., Semboloni E., Van Waerbeke L., Velander M., 2010, A&A, 516, A63 → pages 48, 75

Scranton R., Ménard B., Richards G. T., Nichol R. C., Myers A. D., Jain B., Gray A., Bartelmann M., Brunner R. J., Connolly A. J., Gunn J. E., Sheth R. K., Bahcall N. A., Brinkman J., Loveday J., Schneider D. P., Thakar A., York D. G., 2005, ApJ, 633, 589 → pages 46

Semboloni E., Hoekstra H., Schaye J., 2012, arXiv:1210.7303 → pages 79

Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 2020 → pages 78, 79

Semboloni E., Mellier Y., Van Waerbeke L., Hoekstra H., Tereno I., Benabed K., Gwyn S. D. J., Fu L., Hudson M. J., Maoli R., Parker L. C., 2006, A&A, 452, 51 → pages 48

Simon P., King L. J., Schneider P., 2004, A&A, 417, 873 → pages 70, 80, 83

Simpson F., Heymans C., Parkinson D., Blake C., Kilbinger M., Benjamin J., Erben T., Hildebrandt H., Hoekstra H., Kitching T. D., Mellier Y., Miller L., Van Waerbeke L., Coupon J., Fu L., Harnois-Déraps J., Hudson M. J., Kuijken K., Rowe B., Schrabback T., Semboloni E., Vafaei S., Velander M., 2013, MNRAS, 429, 2249 → pages 50, 81

Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G., Couchman H. M. P., 2003a, MNRAS, 341, 1311 → pages 10, 12, 14

—, 2003b, MNRAS, 341, 1311 → pages 63, 66, 67, 71, 75, 77

SNAP Collaboration: G. Aldering, Althouse W., Amanullah R., Annis J., Astier P., Baltay C., Barrelet E., Basa S., Bebek C., Bergstrom L., Bernstein G., Bester M., Bigelow B., Blandford R., Bohlin R., Bonissent A., Bower C., Brown M., Campbell M., Carithers W., Commins E., Craig W., Day C., DeJongh F., Deustua S., Diehl T., Dodelson S., Ealet A., Ellis R., Emmet W., Fouchez D., Frieman J., Fruchter A., Gerdes D., Gladney L., Goldhaber G., Goobar A., Groom D., Heetderks H., Hoff M., Holland S., Huffer M., Hui L., Huterer D., Jain B., Jelinsky P., Karcher A., Kent S., Kahn S., Kim A., Kolbe W., Krieger B., Kushner G., Kuznetsova N., Lafever R., Lamoureux J., Lampton M., Le Fevre O., Levi M., Limon P., Lin H., Linder E., Loken S., Lorenzon W., Malina R., Marriner J., Marshall P., Massey R., Mazure A., McKay T., McKee S., Miquel R., Morgan N., Mortsell E., Mostek N., Mufson S., Musser J., Nugent P., Oluseyi H., Pain R., Palaio N., Pankow D., Peoples J., Perlmutter S., Prieto E., Rabinowitz D., Refregier A., Rhodes J., Roe N., Rusin D., Scarpine V., Schubnell M., Sholl M., Smadja G., Smith R. M., Smoot G., Snyder J., Spadafora A., Stebbins A., Stoughton C., Szymkowiak A., Tarle G., Taylor K., Tilquin A., Tomasch A., Tucker D., Vincent D., von der Lippe H., Walder J., Wang G., Wester W., 2004, ArXiv Astrophysics e-prints → pages 85

Soucail G., Fort B., Mellier Y., Picat J. P., 1987, A&A, 172, L14 → pages 6

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, Nature, 435, 629 → pages 27

Tyson J. A., Wenk R. A., Valdes F., 1990, ApJ, 349, L1 → pages 6

Van Waerbeke L., 2010, MNRAS, 401, 2093 → pages 27, 44, 47

Van Waerbeke L., Benjamin J., Erben T., Heymans C., Hildebrandt H., Hoekstra H., Kitching T. D., Mellier Y., Miller L., Coupon J., Harnois-Déraps J., Fu L., Hudson M. J., Kilbinger M., Kuijken K., Rowe B. T. P., Schrabback T., Semboloni E., Vafaei S., van Uitert E., Velander M., 2013, ArXiv e-prints:1303.1806 → pages 81, 82

Van Waerbeke L., Bernardeau F., Mellier Y., 1999, A&A, 342, 15 → pages 7

Van Waerbeke L., Mellier Y., 2003, ArXiv Astrophysics e-prints → pages 7, 8

Van Waerbeke L., Mellier Y., Erben T., Cuillandre J. C., Bernardeau F., Maoli R., Bertin E., McCracken H. J., Le Fèvre O., Fort B., Dantel-Fort M., Jain B., Schneider P., 2000, A&A, 358, 30 → pages 7

Van Waerbeke L., Mellier Y., Hoekstra H., 2005, A&A, 429, 75 → pages 48

Van Waerbeke L., Mellier Y., Pelló R., Pen U. L., McCracken H. J., Jain B., 2002, A&A, 393, 369 → pages 48

Van Waerbeke L., White M., Hoekstra H., Heymans C., 2006, Astroparticle Physics, 26, 91 → pages 49

Villumsen J. V., 1996, MNRAS, 281, 369 → pages 7

Wall J. V., Jenkins C. R., 2003, Practical Statistics for Astronomers. Cambridge University Press → pages 54

Walsh D., Carswell R. F., Weymann R. J., 1979, Nature, 279, 381 → pages 5

Weinberg S., 2008, Cosmology. Oxford University Press → pages 2

Wittman D., 2009, ApJ, 700, L174 → pages 50

Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, Nature, 405, 143 → pages 7

Wraith D., Kilbinger M., Benabed K., Cappé O., Cardoso J. F., Fort G., Prunet S., Robert C. P., 2009, PhRvD, 80, 023507 → pages 67

Zhang P., 2010, ApJ, 720, 1090 → pages 15

Zhang P., Pen U., Bernstein G., 2009, ArXiv:0910.4181 → pages 21, 47, 81