Dark Energy properties from the combination of large-scale structure and weak gravitational lensing in the Dark Energy Survey

Carles Sánchez Alonso

Dark Energy properties from the combination of large-scale structure and weak gravitational lensing in the Dark Energy Survey

Author: Carles Sánchez Alonso

Departament de Física
Universitat Autònoma de Barcelona

A thesis submitted for the degree of
Philosophae Doctor (PhD)

Day of defense: 28 September 2017

Director & Tutor:
Dr. Ramon Miquel Pascual
IFAE & ICREA
Edifici Cn, UAB
08193 Bellaterra (Barcelona), Spain
ramon.miquel@ifae.es

CONTENTS

INTRODUCTION

In the past century we have learned a lot about the Universe, and there are two main aspects responsible for this advance. On the one hand, the foundation of General Relativity enabled the development of a coherent, testable theory of the Universe. On the other hand, the improvement of technology and observational techniques have led to the collection of an enormous amount of data. Remarkably, there exists a model for which theory and observations are in *quantitative* agreement. This standard cosmological model, called ΛCDM and presented in Chapter 1 of this thesis, appears to be robust and simple, but only with the addition of two components of unknown nature and origin: dark matter and dark energy. Furthermore, these two components constitute the vast majority of the energy content in the Universe. For that reason, it is of capital importance to understand these components in depth, as they may be holding the key to the discovery of new physics beyond the standard cosmological model and the standard model of particle physics.

Galaxy surveys provide detailed information on the large-scale structure of the Universe, which, in turn, helps us understand its geometry, composition, evolution and fate. Over the past decades, these maps of the Universe have grown from thousands of galaxies in the pioneering Center for Astrophysics (CfA) Redshift Survey (1977-1982) to several million galaxies in the Sloan Digital Sky Survey (SDSS, started in 2000). In this thesis we analyze data from the Dark Energy Survey (DES), which is an ongoing imaging survey, started in 2012, that will cover about one eighth of the sky (5000 sq. deg.) to an unprecedented depth, imaging about 300 million galaxies in 5 broadband filters in the optical and near infrared parts of the electromagnetic spectrum. DES and its necessary associated techniques are presented in detail in Chapter 2, and that, together with the theoretical background introduced in Chapter 1, consititues Part I of this thesis.

The accuracy of the science to be performed with DES and any other photometric galaxy survey strongly depends on the correct estimation of galaxy redshifts, which are related to the distances to those detected galaxies. The photometric redshift estimation technique, to which we devote Part II of this thesis,

relies purely on galaxy colors, and its associated inaccuracies constitute one of the most important sources of observational systematic uncertainties in imaging galaxy surveys and hence in modern cosmology. The principles of this technique are presented, followed by an extensive analysis of different photometric redshift estimation methods applied to DES data, including the characterization of photometric redshift algorithms later used for cosmological studies in DES.

Naturally, galaxy surveys grant us access to the large-scale structure of the Universe through the observation of the galaxy distribution. However, galaxies are not unbiased tracers of the total matter distribution, which appears to be composed predominantly by dark matter. Crucially, imaging galaxy surveys enable the measurement of the weak gravitational lensing effect, which produces small distortions in the shapes of distant galaxies due to the gravitational pull of foreground structures, and is sensitive to the total, both luminous and dark, matter distribution. The combination of galaxy clustering and weak gravitational lensing, which is the subject of Part III of this thesis, enables the characterization of the galaxy–dark matter connection and unlocks crucial information about the growth of cosmic structures and the local expansion rate, representing one of the more promising avenues to understand the underlying physical mechanism responsible for the accelerated expansion of the Universe. In this part we present a study of potential systematic effects that can plague the galaxy clustering and weak gravitational lensing measurements, and study their combination to obtain cosmological constraints.

Beyond the usage of galaxies to learn about dark energy and cosmology, there is important and independent information in studying cosmic voids. Cosmic voids are large, underdense regions in space surrounded by a filamentary network of galaxies and dark matter. They are usually identified in spectroscopic galaxy surveys, where accurate 3D large-scale structure information is available. Nevertheless, with the enormous amount of data coming from current and future imaging surveys, developing void science for this kind of surveys has the potential to exploit the constraining power of cosmic voids and represent a key advance for cosmology. That is the main subject of Part IV and Chapter 5 of this thesis, where we present a new void finding algorithm designed for photometric surveys and demonstrate the emptiness of the resulting set of voids measuring the gravitational lensing signal around them.

Finally, at the end we present the summary and major conclusions from the thesis, also analyzing its implications and potential extensions.

Part I

PRELIMINARS

Chapter 1

COSMOLOGICAL FRAMEWORK

Cosmology is the branch of physics that studies the origin, evolution and structure of the Universe as a whole, and it was mainly developed in the twentieth century, definitely moving from philosophy to science. Modern cosmology is based on the belief that the place we occupy in the universe is not special, statement that we know as the *cosmological principle*. In particular, it states that the universe is homogeneous[1] and isotropic[2]. Needless to say, the *cosmological principle* is not exact, but an approximation that holds better and better the larger the length scales we consider.

Before 1900 the universe was considered a static and inalterable system, due to the appearance of the dark sky and all far-away celestial objects occupying fixed angular positions on it. Together with the basis established by the *cosmological principle*, the evolution of the universe as a whole should be governed by the laws of gravity, as the other fundamental forces do not play a role on such large scales. In 1915, Einstein presented his theory of General Relativity (GR, 1915), the most accurate description of gravity, represented in the theory as the consequence of the curvature of spacetime produced by the distribution of mass and energy in the universe.

Shortly after these theoretical developments, in 1929, Edwin Hubble measured the motion of various galaxies along the line of sight using the 100-inch Hooker telescope at Mount Wilson Observatory and he found that most of these galaxies were indeed receding, the faster the further they were from us. It was therefore realized that the universe was actually not static, but expanding. This observational breakthrough eventually led to the development of the Big Bang theory, which describes the Universe as expanding from an initial very high density and high temperature state. The Big Bang theory still remains the basis of today's

[1] The same in all locations.
[2] The same in all directions.

Standard Cosmological Model, supported by other crucial observations like the abundance of primordial elements and the measurements of the relic light from the early Universe, the Cosmic Microwave Background radiation.

Within the framework of GR, which naturally accomodates (or even predicts!) a dynamical Universe, the expansion rate of the Universe depends on its energy content in a way that, for instance, a universe containing only matter should eventually slow down due to the attractive force of gravity. However, in 1998, observations of type Ia supernovae (SNe) at distances up to about 6 bilion light years by two independent research groups, led by Saul Perlmutter and by Brian Schmidt and Adam Riess respectively, revealed that presently the expansion is instead accelerating.

This acceleration of the expansion is attributed in the Standard Cosmological Model to the presence of an unknown kind of fluid, possibly related to a cosmological constant, that we call Dark Energy (DE). If $p = w\rho$ is the equation of state relating pressure p to energy density ρ (and being w the equation-of-state parameter), then the requirement for that fluid is to have $w < -1/3$, representing a negative pressure. So far, we do not know what is the nature of this energy content that makes the expansion of the universe to accelerate. Moreover, current measurements of the energy content of the universe show that DE accounts for about 70% of the total energy density of the universe. Of the remainder, more than 25% is due to an unknown form of matter (called Dark Matter, DM) and only less than 5% of the energy density corresponds to ordinary matter like protons, electrons and neutrinos.

1.1 THE SMOOTH UNIVERSE

1.1.1 THE FIELD EQUATIONS

As mentioned above, gravity is the fundamental force describing the dynamics of bodies on large, cosmological scales. Over such large distances, space itself behaves differently from the well-known Euclidean space we are used to in everyday life. In that way, gravity is described by the theory of General Relativity, which relates the mass-energy content with the geometry of space-time through the Einstein's field equations (in natural units[3]):

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.1}$$

[3]Natural units are defined by normalizing some universal physical constants to one. We choose $c = 1$ (speed of light), $\hbar = 1$ (Planck constant) and $k_B = 1$ (Boltzmann constant). We will use this system of units from now on, unless we state otherwise.

where $G_{\mu\nu}$ is the Einstein tensor depending on space-time second order derivatives of the metric $g_{\mu\nu}$ and $T_{\mu\nu}$ is the *energy-momentum* tensor depending on the mass-energy distribution.

1.1.2 THE FLRW METRIC

The *cosmological principle* presented at the beginning of this chapter states that the universe is homogeneous and isotropic. These properties are fundamental in order to find a particular simple solution to the field equations (1.1).

We start working on that solution by building a metric from the above mentioned principle. The metric of space-time used in Cosmology comes from the *Minkowski* metric used in Special Relativity,

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + dl^2 = -dt^2 + dr^2 + r^2 d\Omega^2, \tag{1.2}$$

adding some assumptions (with $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$, in spherical coordinates, being used). For a space-time that is homogeneously expanding (or contracting), the spatial component of the metric has to allow for time dependence. For that purpose, it is convenient to define a coordinate system that expands (or contracts) with the universe (see Fig. 1.1). The distance between two galaxies D_{12}, which changes with time in a non-static universe, can be parametrized in such coordinate system by a comoving distance χ_{12} (independent of time) times a scale factor $a(t)$ as

$$D_{12}(t) = a(t)\chi_{12}, \tag{1.3}$$

where $a(t)$ is normalized to the present value $a(t_0) = 1$. Allowing also for some curvature in space, the only metric that is compatible with homogeneity and isotropy in the universe can be shown to be the so-called *Friedmann-Lemâitre-Robertson-Walker (FLRW)* metric:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right], \tag{1.4}$$

where k is an undetermined constant that represents the curvature of space-time. The possibilities k positive, zero or negative correspond to the three possible spatial geometries spherical, flat or hyperbolic, respectively.

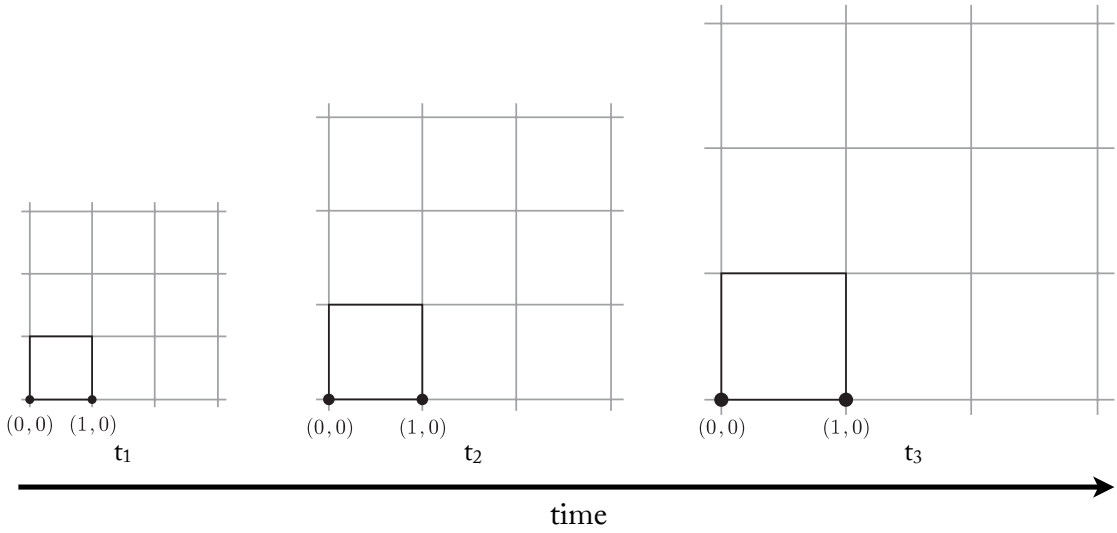

Figure 1.1: The comoving coordinate system. As the universe expands, the comoving distance between reference points on this imaginary coordinate grid remains constant. In turn, the physical distance is given by the comoving distance times the scale factor $a(t)$. As seen in the figure, for $t_3 > t_2 > t_1$ we have $a(t_3) > a(t_2) > a(t_1)$ and hence the physical distance gets larger as time evolves in an expanding universe.

1.1.3 THE FRIEDMANN EQUATIONS

The main possible constituents of the universe are all examples of so-called perfect fluids, and are described by the following energy momentum tensor:

$$T_{\mu\nu} = \text{diag}(\rho, p, p, p), \tag{1.5}$$

ρ being the energy density and p the pressure. Given that energy-momentum tensor form, if we insert the metric (1.4) into (1.1) we obtain the famous Friedmann equations, describing the time evolution of the scale factor:

$$H^2(t) \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{1.6}$$

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{3}(\rho + 3p). \tag{1.7}$$

$H(t) = \frac{\dot{a}}{a}$ is known as the Hubble parameter and it measures the expansion rate of the universe. Its present value, $H_0 = H(t_0)$, is usually parametrized in the form $H_0 = 100\frac{\text{km}}{\text{Mpc·s}} \cdot h$, with h a dimensionless number.

From the two last equations we can build a third one, that is not independent

but useful, called the fluid equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0. \tag{1.8}$$

With the Friedmann equations in hand, one can solve for the time evolution of different components of the universe, given their equation of state $p = p(\rho)$:

- matter: $p_m = 0 \Rightarrow \rho_m \propto a^{-3}$;
- radiation: $p_r = \rho_r/3 \Rightarrow \rho_r \propto a^{-4}$;
- cosmological constant: $\rho_\Lambda = \text{constant} \Rightarrow p_\Lambda = -\rho_\Lambda$;
- general DE: $p_{DE} = w\,\rho_{DE} \Rightarrow \rho_{DE} \propto a^{-3(1+w)}$, if w is constant;

where we have used the fluid equation.

The densities ρ of all components can be normalized to the critical density ρ_{cr}, the one that makes the universe flat today, from (1.6): $\rho_{cr} = 3H_0^2/8\pi G$. With this normalization, the densities of all components are written as $\Omega_i = \rho_i/\rho_{cr}$. The first Friedmann equation is usually written in terms of all these normalized density parameters (simply called cosmological parameters):

$$H^2(a) = H_0^2\left[\Omega_r\, a^{-4} + \Omega_m\, a^{-3} + (1-\Omega_0)\, a^{-2} + \Omega_\Lambda\right], \tag{1.9}$$

where $\Omega_0 = \Omega_r + \Omega_m + \Omega_\Lambda$ is the total density today. Our knowledge about these parameters has been significantly improved over the last decade, in the sense that different kinds of observations now yield concordant values and errors of a few percent in their determination. Although they are not exactly the best-fit parameters, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $\Omega_r = 0$ are good approximations.

1.2 FUNDAMENTAL OBSERVATIONS

1.2.1 THE EXPANSION OF THE UNIVERSE: HUBBLE'S LAW

We have already anticipated at the beginning of the chapter that Hubble found evidence for the expansion of the universe by measuring the movement of distant galaxies along the line of sight. But he actually did not measure velocities directly, he instead observed that the light from a given galaxy was shifted further toward the red end of the spectrum the further that galaxy was from our galaxy. At low redshifts (low recession velocities), the relation between *redshift* (z) and velocity

(υ) is given by the well known Doppler effect:

$$z = \frac{\lambda_o}{\lambda_e} - 1 = \sqrt{\frac{1+\upsilon/c}{1-\upsilon/c}} - 1 \simeq \frac{\upsilon}{c}, \tag{1.10}$$

with subscript '*e*' denoting quatities at the emission of the photons, and '*o*' denoting quantities at the observation.

Back in 1929, Hubble not only noted that distant galaxies are receding from us, he also found that the recession velocity increases linearly with distance. This behavior is exactly what one expects in an expanding universe. We have seen that the physical distance between two galaxies is given by $D = a\chi$, where χ is the comoving distance (a distance measured in the coordinate system that expands with the universe). In the absence of comoving motions ($\dot{\chi} = 0$), the relative velocity between these two galaxies is:

$$\upsilon = \dot{D} = \dot{a}\chi = \frac{\dot{a}}{a}D = HD. \tag{1.11}$$

Then, at low redshift where (1.10) is valid, there exists a linear relation between redshift and distance. This relation is exactly what Hubble found (Fig. 1.2), since he was looking at galaxies not very far from us, with small redshifts. There are two main reasons for the dispersion in Fig 1.2. On the one hand, since the galaxies considered are not very far from us, they have velocities driven by local gravitational fields (comoving motions) that are not due to the expansion of the universe but are comparable in speed. These velocities, called *peculiar* velocities, introduce a dispersion. On the other hand, the estimation of the distance to those galaxies by Hubble was not very accurate, introducing further dispersion.

The redshift z can be related to the scale factor a. Let's consider two wavefronts emitted at times t_e and $t_e + \lambda_e/c$. Taking into account that light travels in geodesics ($ds^2 = 0$, in (1.4)), we have:

$$\int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_R^0 \frac{dr}{\sqrt{1-kr^2}} \;\; ; \;\; \int_{t_e+\lambda_e/c}^{t_0+\lambda_0/c} \frac{dt}{a(t)} = \int_R^0 \frac{dr}{\sqrt{1-kr^2}}. \tag{1.12}$$

The right-hand sides of the two integral equations above are identical, so we can write:

$$\int_{t_e}^{t_0} \frac{dt}{a(t)} = \int_{t_e+\lambda_e/c}^{t_0+\lambda_0/c} \frac{dt}{a(t)} \;\Rightarrow\; \int_{t_e}^{t_e+\lambda_e/c} \frac{dt}{a(t)} = \int_{t_0}^{t_0+\lambda_0/c} \frac{dt}{a(t)}. \tag{1.13}$$

Now we can consider that the scale factor is constant over the small period of one

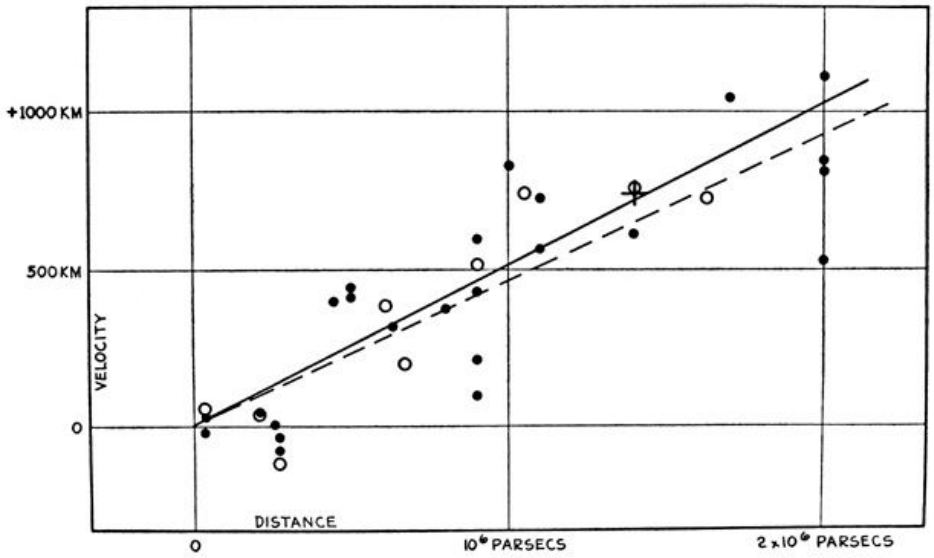

Figure 1.2: Original plot from Hubble. Note the wrong units (km instead of km/s for velocity) in the vertical axis.

cycle of a light wave, and therefore:

$$\frac{t_e + \lambda_e/c}{a_e} - \frac{t_e}{a_e} = \frac{t_0 + \lambda_0/c}{a_0} - \frac{t_0}{a_0} \;\Rightarrow\; \frac{\lambda_0}{\lambda_e} = \frac{a_0}{a_e}. \tag{1.14}$$

Using the definition of redshift in Eq. (1.10) we find the relation between redshift and the scale factor to be:

$$1 + z = \frac{a_0}{a_e} = \frac{1}{a_e}. \tag{1.15}$$

1.2.2 THE COSMIC MICROWAVE BACKGROUND

After the observation of an expanding universe, cosmologists could reach the conclusion that the Universe was smaller in size in the past. This means, for instance, that earlier in time the physical distance between us and other galaxies was smaller than it is today. It also means that the electromagnetic radiation filling the Universe was more energetic (or hotter) than it is today, as radiation energy is inversely proportional to its wavelength and hence to the scale factor, and that is smaller for earlier times in an expanding universe. If we define the temperature of that radiation to be T_0 today, then we can simply express the temperature at any time as

$$T(t) = \frac{T_0}{a(t)} \tag{1.16}$$

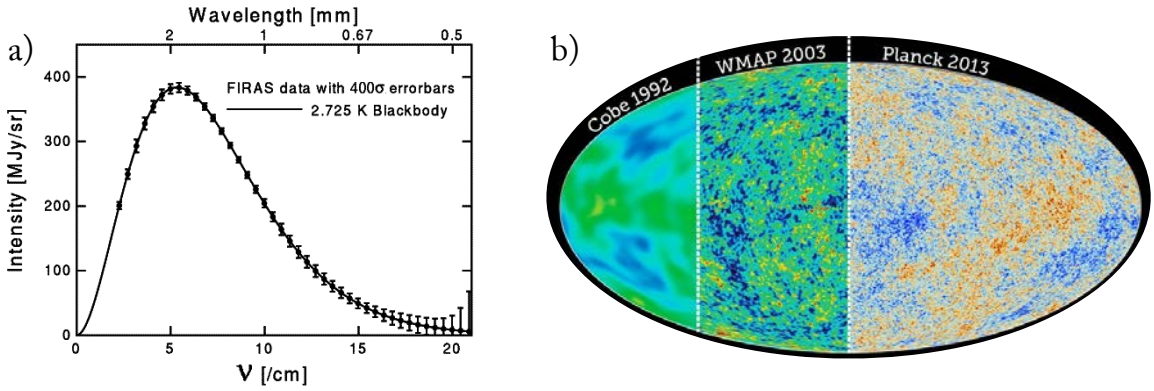

Figure 1.3: a) Intensity of CMB radiation as a function of frequency (and wavelength) for the COBE FIRAS experiment (Mather et al 1994), showing impressive agreement with a blackbody sepctrum (note that errorbars are magnified by a factor of 400 in the plot!). b) Anisotropies in the CMB (color scale shows relatve differences of order 10^{-5}) as seen by three different satellites: COBE, WMAP and Planck; clearly apparent is the increase in angular resolution from one experiment to another.

where $a(t)$ today is simply $a_0 = 1$. Therefore, by deduction, the temperature of the Universe must have been very high at early times. In particular, it may have been high enough to be able to ionize atoms such as hydrogen or Helium, and then the Universe at that time must have been a hot ionized plasma of particles and radiation (and dark matter and dark energy). In that state, photons and electrons were coupled via Thomson scattering, and the mean free path of photons was very short, effectively making the Universe *opaque*. Also, any hydrogen atoms being formed in that period were quickly ionized by ambient photons. Crucially, as a consequence of these constant interactions, photons were in equilibrium in that early period of the Universe, and radiation in equilibrium has a very particular spectral distribution: the blackbody spectrum.

As the Universe expanded and cooled, the energy of ambient photons was no longer high enough to ionize Hydrogen, and the mean free path of photons grew drastically, effectively making the Universe *transparent*. In the mid twentieth century, cosmologists started looking for the remnants of this radiation in equilibrium coming from the very early Universe, as a proof of the Big Bang theory and a way to see how the Universe looked like when it was in its infancy. The Cosmic Microwave Background (CMB), called like this for its higher intensity in the microwave part of the radio spectrum today, was first detected by Penzias & Wilson (1965), in a single-wavelength experiment. This first detection was further confirmed by other experiments but it was the COBE satellite, in 1992, that impressively extended it to a broad range of wavelengths, confirming its agreement with a blackbody spectrum and hence providing direct observational evidence for the Big Bang (Panel a) in Fig. 1.3).

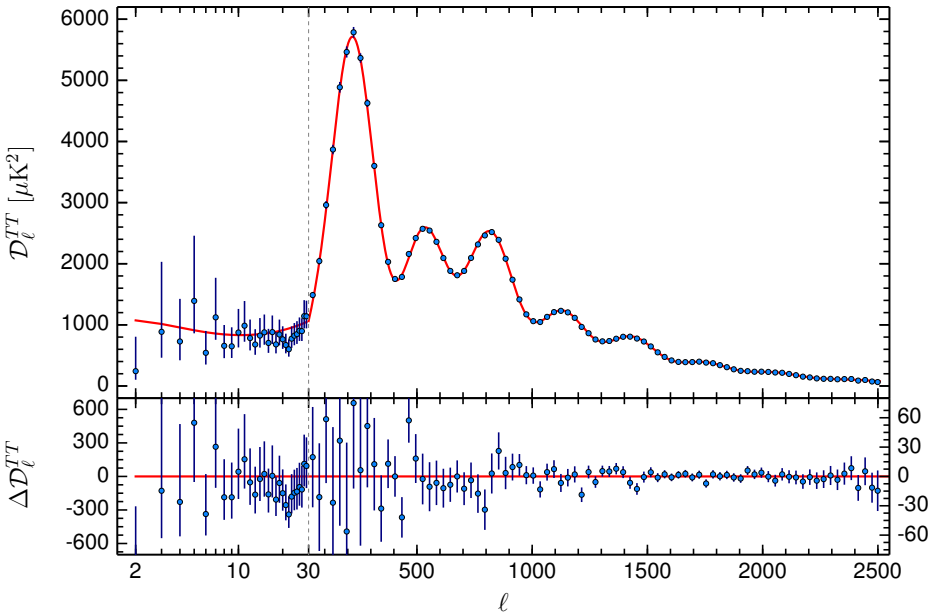

Figure 1.4: Planck 2015 CMB power spectrum (blue dots and error bars), with $D_l \equiv l(l+1)/(2\pi)C_l$, and ΛCDM best-fit model (red line). The upper panel shows the spectrum and the lower panel the residuals. Note the change of scale from logarithmic to linear at the vertical dotted line, $l = 29$, and the change in the residuals y-axis between these two regimes.

After COBE, many experiments including other satellites like WMAP and more recently Planck have measured and characterized the CMB properties. In detail, now we know that CMB radiation comes from the Universe when it was about 400,000 years old, (redshift $z \simeq 1100$) and its spectrum today corresponds to a blackbody at a tempreature of $T_0 = 2.72548 \pm 0.00057$ K (Fixsen 2009). Even though the CMB represents the most precise blackbody spectrum in nature, and it looks the same in all directions without variations at the percent level (what is known as the CMB monopole), it is not perfectly isotropic. First, there exists an anisotropy of the order of 1 part in 10^3 which corresponds to the Doppler effect caused by the movement of the Milky Way with respect to the CMB reference rest frame; this is known as the CMB dipole anisotropy. More interestingly, once the dipole is corrected for, smaller temperature variations at the 1 part in 10^5 level also exist and have been studied in detail (Panel b) in Fig. 1.3). Such anisotropies are caused by the density fluctuations in the early Universe which eventually led to the structures we see today (galaxies, clusters of galaxies, filaments...). In particular, the most important effect comes from oscillations of the baryon-photon plasma in the early Universe. Gravity tries to compress the photon-baryon fluid sitting on primordial gravitational potential wells, while radiation pressure acts in the opposite way, resulting in acoustic oscillations. These anisotropies are usually

analyzed using a decomposition in spherical harmonics $Y_{l,m}(\theta,\phi)$ such as:

$$\frac{\Delta T(\theta,\phi)}{T_0} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{l,m} Y_{l,m}(\theta,\phi), \tag{1.17}$$

where the $a_{l,m}$ coefficients, provided they are independent, can be completely characterized by what is known as the CMB power spectrum:

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^{l} \left|a_{l,m}\right|^2, \tag{1.18}$$

which describes CMB fluctuations at a given multipole moment, l, corresponding to an angular scale of about $l \sim \pi/\theta$. Figure 1.4 shows the Planck 2015 measured CMB power spectrum and its excellent agreement with the ΛCDM best-fit theoretical model. The baryon-photon acoustic oscillations are responsible for its peak structure, as photons decoupling when a particular mode is in its peak amplitude create ressonances which correspond to the different peaks. There are other effects occurring at later times in the Universe which are also imprinted in the CMB, as its photons are free to interact with ionized electrons and they also feel the gravitational fields of the late-time large-scale structure of the Universe. As a consequence of the variety of the physical phenomena involved, the analysis of the CMB power spectrum provides an enormous amount of information about the composition of the Universe and hence about the cosmological parameters defined in Eq. (1.9). For instance, the position and amplitude of the first peak with respect to the secondary peaks provides crucial information about the geometry or curvature of the Universe, and about the matter and baryon densities. In fact, some of the tighest constraints on cosmological parameters come nowadays from the analysis of the latest CMB experiments such as the Planck satellite, and they are also crucial for the combination with different cosmological probes, especially those probing the late-time Universe.

1.2.3 THE ABUNDANCE OF PRIMORDIAL ELEMENTS

As another consequence of Eq. (1.16), the temperature in the Universe must have been high enough for nuclear reactions to take place, as it happens, for instance, in the interiors of stars such as the Sun. These nuclear reactions may have allowed some light elements, like Helium or Lithium, to form from the combination of protons and neutrons during that period of the early Universe. This process is refererd to as Big Bang Nucleosynthesis (BBN).

The Big Bang model predicts the abundances of light elements formed from

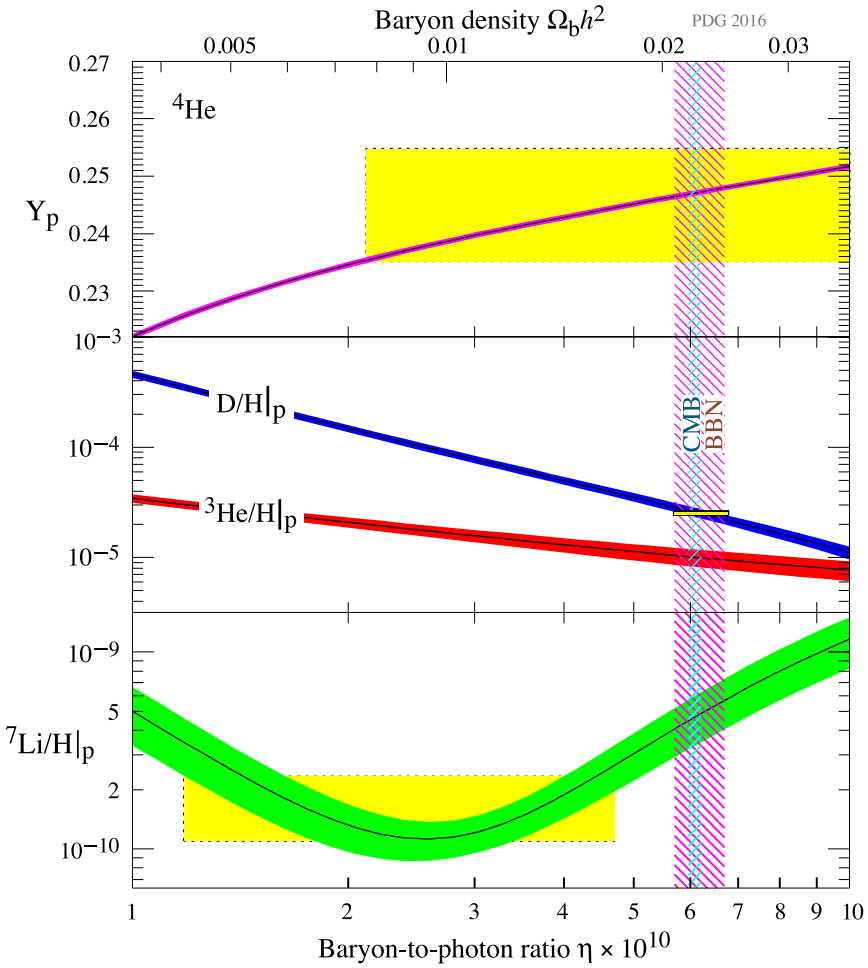

Figure 1.5: From PDG Astrophysics and Cosmology Review (2015): Lines and their bands in this plot show the Big Bang Nucleosynthesis predicted 95% CL abundances of primordial ^{4}He, D, ^{3}He and ^{7}Li as a function of the baryon-to-photo ratio. Yellow boxes indicate the corresponding observations. The vertical narrow band corresponds to the CMB measurement of the baryon-to-photon ratio and the wider band shows the constraints on same quantity from the combination of the different abundance measurements, both at 95% CL.

that process as a function of the fraction of ordinary matter present in the early Universe. The latter can be well determined by analyzing CMB data from experiments such as WMAP or Planck. Using these data, the predictions for the abundances of light elements are in good overall agreement with observations (Fig. 1.5). This is a remarkable validation of the Big Bang theory, especially given the fact that these abundances span about nine orders of magnitude.

1.3 DISTANCES IN THE UNIVERSE

We have seen in §1.2 that redshift and distance are directly related. However, the definition of distance in Cosmology is quite ambiguous, since the distance to a given object varies with the propagation time of the emitted light, due to the expansion of the universe. That's why there exist several definitions of distance, that are used for different applications.

1.3.1 COMOVING DISTANCE

The *comoving distance* measures the distance between two objects in a coordinate system that expands with the universe. In this way, if both objects have no peculiar velocities then that distance remains constant.

For radial light rays ($d\theta = 0$ and $d\phi = 0$), when traveling along null geodesics ($ds^2 = 0$), the metric reduces to:

$$dt^2 = a^2(t)\frac{dr^2}{1-kr^2} \equiv a^2(t)d\chi^2. \tag{1.19}$$

And then we can define the *comoving distance* as the following integral:

$$\chi = \int_{t_e}^{t_0} \frac{dt'}{a(t')} = \int_0^z \frac{dz'}{H(z')}, \tag{1.20}$$

where in the second equality we have used the definiton of H in Eq. (1.6) and the relation between redshift and the scale factor in Eq. (1.15) as:

$$\frac{dz}{dt} = \frac{-1}{a^2}\frac{da}{dt} = \frac{-\dot{a}}{a^2} = \frac{-H}{a} \Rightarrow \frac{dt}{a} = -\frac{dz}{H} \tag{1.21}$$

1.3.2 ANGULAR DIAMETER DISTANCE

One way to measure distances in astronomy is to measure the angle θ subtended by an object of already known physical size l. The distance to that object

is called the *angular diameter distance* D_A, and is defined (assuming θ is small) as:

$$D_A = \frac{l}{\theta}. \tag{1.22}$$

In an expanding universe, the light we observe from the object was emmited when the Universe had the scale factor a, and hence the comoving size of the object is l/a. Since the comoving distance to the object is χ, the angle subtended is $\theta = (l/a)/\chi$. Now, using (1.22) we can write D_A as:

$$D_A = a\chi = \frac{\chi}{1+z}. \tag{1.23}$$

1.3.3 LUMINOSITY DISTANCE

The *luminosity distance* D_L is defined through a relationship between the bolometric[4] luminosity L of an object and its bolometric flux F:

$$F = \frac{L}{4\pi D_L^2}. \tag{1.24}$$

Working with comoving coordinates, centering this time the source at the origin, the observed flux is written as:

$$F = \frac{L(\chi)}{4\pi\chi^2}, \tag{1.25}$$

where $L(\chi)$ is the luminosity through a comoving spherical shell of radius χ. If we assume, for simplicity, that all photons are emitted at the same energy, then $L(\chi)$ is the energy of the photons times the number of photons crossing a comoving spherical shell per unit time. In a given time interval, photons travel farther on the comoving grid at early times than at late times, since the corresponding physical distance at early times is smaller. Then, the number of photons passing through a shell in the given time interval is smaller today than at emission, by a factor of $a = 1/(1+z)$. In a similar way, the energy of the photons is smaller today than at emission (also by a factor of a) due to the increase of the associated wavelength driven by the expansion. Therefore, the total energy per unit time crossing a comoving spherical shell of radius χ (being χ the comoving distance between us and the source) is smaller than the luminosity at the source by a factor of a^2. The observed flux, then, is:

$$F = \frac{La^2}{4\pi\chi^2}. \tag{1.26}$$

[4]Integrated over all frequencies.

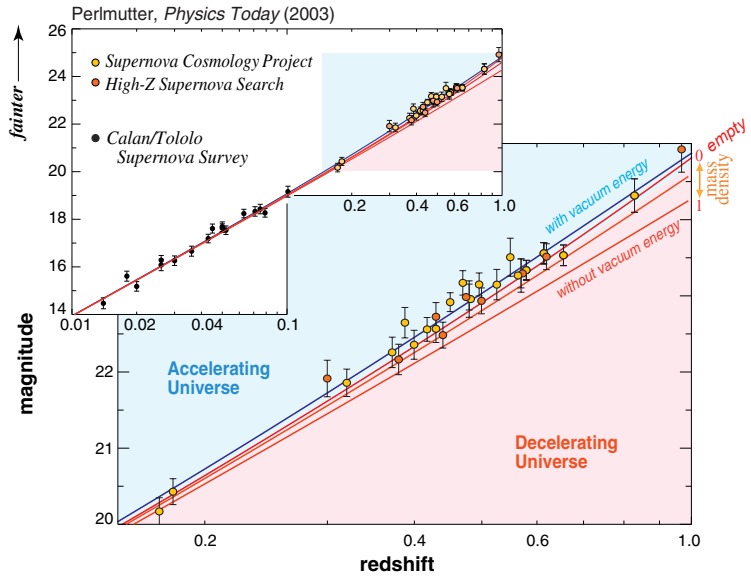

Figure 1.6: Plot showing the luminosity distance, scaled in magnitudes $m \propto \log(D_L)$, as a function of redshift (1.28). It contains data from the two groups that discovered the acceleration of the universe. One can see that Hubble's law starts to fail beyond $z = 0.1$, diverging differently depending on the assumed cosmic densities of mass and DE. Different red lines represent different models of DE and mass densities changing from ρ_{cr} to zero. The blue line shows the best fit, corresponding to a mass density of about $\rho_{cr}/3$ and a DE density twice that.

Finally, comparing this to (1.24) we obtain the *luminosity distance*:

$$D_L = \frac{\chi}{a} = (1+z)\chi = (1+z)^2 D_A. \tag{1.27}$$

1.4 THE ACCELERATING UNIVERSE

In the late 1990's two independent teams, the Supernova Cosmology Project and the High-z SN Search, took advantage of a couple of developments in the field. One was the realization that type Ia supernovae (SNe) can be turned into standard candles[5], and the other was the development of large CCD cameras to be used in ground-based telescopes with the capability of performing systematic searches of these supernovae by comparing deep, wide images taken weeks apart. With these developments in hand, they measured the generalized Hubble relationship,

$$D_L = (1+z)\chi = (1+z)\int_0^z \frac{dz'}{H(z')}, \tag{1.28}$$

[5]In astronomy, a standard candle is a type of astrophysical objects which have definite luminosity due to some characteristic quality possessed by the entire class of objects.

to much larger distances than it had been measured before, and they found that distant SNe were fainter than expected in a decelerating universe (note that the cosmological parameters enter in (1.28) through $H(z')$). Overall, both groups provided evidence for $\Omega_\Lambda > 0$ at greater than 99% confidence level. Figure 1.6 shows the data corresponding to that discovery, and also the best fit corresponding to cosmological parameters about $\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$ and $\Omega_r = 0$. This cosmological model is known as the ΛCDM model.

1.5 THE LARGE-SCALE STRUCTURE OF THE UNIVERSE

At the end of §1.2.2 and Fig. 1.3 we have seen the CMB is not perfectly uniform, but there are small anisotropies imprinted on it as measured by several experiments. Actually, a completely homogeneous density field cannot exist: quantum fluctuations must occur on microscopic scales. However, the fluctuations observed in the CMB are way too large compared to those expected from a pure quantum process.

The inflationary theory, which describes a period of exponential expansion of space in the early Universe, was initially proposed as a way to solve the flatness and the horizon problems. These refer, respectively, to the fact that Universe appears to be flat, and also homogeneous and isotropic while there being widely separated regions of the observable Universe that cannot have equilibrated because, due to the expansion, they move apart from each other faster than the speed of light and then have never been in causal contact. Importantly, it also provides a simple mechanism for the generation and growth of perturbations: quantum fluctuations in the very early Universe were magnified to cosmic size due to the exponential expansion, becoming the seeds of the anisotropies observed in the CMB and, at later times, the Large-Scale Structure (LSS) of the Universe.

1.5.1 LINEAR STRUCTURE FORMATION

The initial conditions set by the inflationary theory determine how the matter distribution looks like in the present-day Universe, provided we assume a model for the evolution of structure. On large scales, perturbations remain small and linear theory can be used. Therefore, by analyzing the LSS matter distribution as measured by galaxy surveys, which is the main subject of this thesis, we can derive cosmological parameters.

For that purpose, let us define the mean matter density of the Universe as $\bar{\rho}$, and the local density at any three-dimensional position as $\rho(\mathbf{x})$. Then we can

define the matter density fluctuations as:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}. \tag{1.29}$$

If δ is positive (negative) at some position $\mathbf{x}$, this means there is a local matter over(under)-density in there. Following a Newtonian approach, valid if the length scales of these perturbations are smaller than the cosmological horizon, $d_H = c/H_0$, and assuming the perturbations are small, $\delta \ll 1$, one can combine mass and momentum conservation with the Poisson equation for the Newtonian gravitational potential and show that the evolution of density perturbations should be described by the following second order harmonic differential equation:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}H^2\delta = 0, \tag{1.30}$$

where H is the Hubble parameter defined in Eq. (1.6). This equation has a general solution consisting on two independent modes, growing and decaying, for the evolution of perturbations. Since the growing mode dominates, we write the solution as:

$$\delta(\mathbf{x}, t) = \frac{D(t)}{D(t_0)}\delta(\mathbf{x}, t_0), \tag{1.31}$$

where $D(t)$ is called the growth factor, describing the growth of structures in the expanding Universe. $D(t)$ depends on the cosmological density paramters and can be computed explicitly for different cosmologies. For instance, $D(t) = a(t)$ for the case $\Omega_\Lambda = 0$, $\Omega_m = 1$, and $D(t) < a(t)$ for the case of a ΛCDM cosmology.

1.5.2 THE PRIMORDIAL POWER SPECTRUM

A successful cosmological theory of structure formation cannot predict the exact distribution of matter in the Universe. Instead, it should predict the statistical properties of such distribution, in particular concerning how it deviates from homogeneity, what we know as the clustering in a distribution. For a density field $\delta(\mathbf{x})$ that is homogeneous and isotropic on large scales, we can define the 2-point correlation function (2PCF) as

$$\xi(|\mathbf{x} - \mathbf{y}|) = \xi(r) = \langle \delta(\mathbf{x})\delta^*(\mathbf{y}) \rangle, \tag{1.32}$$

where the brackets $\langle ... \rangle$ denote an ensemble average over many realizations or, lacking an ensemble of universes, an average over different volumes, assuming that well-separated regions of the Universe can be considered as independent realizations of the same underlying density field. Implicit in the definition above

we have included homogeneity, ξ depends only on $(\mathbf{x}-\mathbf{y})$ and not $(\mathbf{x},\mathbf{y})$ and, additionally, isotropy, as ξ depends only on $r = |\mathbf{x}-\mathbf{y}|$. It is also useful to define the Fourier transform of the 2PCF, known as the power spectrum:

$$P(|\mathbf{k}|) = \int d^3x e^{-i\mathbf{x}\cdot\mathbf{k}} \xi(|\mathbf{x}|), \tag{1.33}$$

which obeys:

$$\left\langle \delta(\mathbf{k})\delta^*(\mathbf{k}')\right\rangle = (2\pi)^3 \delta_D(\mathbf{k}-\mathbf{k}')P(|\mathbf{k}|), \tag{1.34}$$

where δ_D is the three-dimensional Dirac delta function and $\delta(\mathbf{k})$ is can be expressed as:

$$\delta(\mathbf{k}) = \int d^3x e^{i\mathbf{x}\cdot\mathbf{k}} \delta(\mathbf{x}). \tag{1.35}$$

The inflationary theory predicts that the primordial power spectrum $P_0(k)$ should not depend on a characteristic scale, and then it should have the form of a power law:

$$P_0(k) \propto k^{n_s}. \tag{1.36}$$

The case where $n_s = 1$ is often refereed to as the Harrison-Zeldovich spectrum (Harrison 1970, Zeldovich 1972). In fact, inflation predicts a spectral index that is smaller but very close to unity. With the latest measurements using the Planck satellite, $n_s = 0.9655 \pm 0.0062$ (TT + LowP, Planck2015 et al).

The amplitude of the power spectrum cannot be deduced from theory but should be measured from observations. This amplitude is usually parameterized in terms of σ_8, defined as the square root of the variance of the density fluctuations in spheres of radius 8 h^{-1} Mpc scattered througout the Universe:

$$\sigma_R^2 = \frac{1}{2\pi^2} \int k^2 P(k) \left[\frac{3j_1(kR)}{kR}\right]^2 dk, \tag{1.37}$$

where $R = 8\,h^{-1}$ Mpc, and we have integrated the power spectrum using a window function which is the Fourier transform of a spherical top-hat of radius R: $W(kR) = 3j_1(kR)/(kR)$, with $j_1(kR)$ being the spherical Bessel function of order 1.

The matter power spectrum today is usually connected to the primordial power spectrum via the so-called transfer function, $T(k,z)$, which depends on the contents of the Universe and hence is a function of the particular cosmological model:

$$P(k,z) = T(k,z)^2 P_0(k). \tag{1.38}$$

The redshift dependent part of the transfer function can be expressed as function

of the scale and the growth factors:

$$T(k,z) = \frac{D(z)}{D(z=0)} a(z) T(k). \tag{1.39}$$

For ΛCDM, there exist analytic fitting formulas for the tranfer function. As an example, particularly popular is the one provided by Bardeen, Bond, Kaiser & Szalay (BBKS, 1996):

$$T(q) = \frac{\ln[1+2.34q]}{2.34q} \left[1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4\right]^{-1/4}, \tag{1.40}$$

where $q = k/(\Gamma\, h\ \mathrm{Mpc}^{-1})$. $\Gamma \equiv \Omega_m h$ is called the shape parameter and can also depend weakly on the baryon density, although that dependency is usually neglected. Using this transfer function, the resulting power spectrum predicts the formation of small structures first, merging to form bigger structures later in time, as we shall discuss next.

1.5.3 NON-LINEAR EVOLUTION

Structures in the Universe grow under the influence of gravity, which attracts matter around overdensities thereby making the density field less and less homogeneous as time passes. In ΛCDM, the formation of structures is hierarchical: small structures form first around peaks in the initial density field, collapsing into virialized dark matter halos, and larger structures form later, either from the accretion of mass in the neighborhood of existing overdensities or from the merging of smaller halos. Following the Spherical Collapse model (Mo & White 1996, Peebles 1980, Press and Schechter 1974), overdensities collapse to form a gravitationally bound dark matter halo when they exceed a given value for the critical density δ_c. Under this paradigm, baryonic matter falls into already formed dark matter halos by feeling their gravitational attraction, as the temperature of baryonic matter was still too high to form bound objects from their own gravitational field, as opposed to cold dark matter (CDM) which could form gravitationally bound structures earlier in time. The accretion of baryonic matter by dark matter halos sets the (gravitational) conditions for galaxy formation to happen.

In order to study the evolution of perturbations that become the seeds of galaxies, linear theory is no longer valid and a different approach is needed. The non-linear collapse of dark matter halos, the possible merging between them and the subsequent galaxy formation processes are so complicated that nowadays most predictions about the non-linear power spectrum and hence about the clustering of matter on small scales are the result of large N-body simulations. The matter

structures generated in the simulations are then used to obtain fitting formulas for the non-linear power spectrum, as in e. g. Smith et al. (2003) and more recently Takahashi et al. (2012).

1.5.4 THE GALAXY 2-POINT CORRELATION FUNCTION

Since galaxies are the main observable of the Large-Scale Structure of the Universe, next we will show how to study the properties of their distribution in a statistical sense, which enables the comparison between theory and observations. The clustering in the distribution of galaxies can be quatified by measuring the 2-point correlation function (2PCF) of galaxy positions, $\xi_g(r)$, which quantifies the departure of the galaxy distribution from homogeneity. An alternative is to measure the Fourier transform of the galaxy 2PCF, the galaxy power spectrum. The galaxy 2PCF, $\xi_g(r)$, is defined as the excess probability dP of finding a galaxy in a volume element dV at a 3D separation r from another galaxy above a uniform Poisson distribution,

$$dP = \bar{n}(1 + \xi_g(r))dV, \tag{1.41}$$

where $\bar{n}$ is the mean galaxy density. In practice, $\xi_g(r)$ is estimated from counting pairs of galaxies and random points following the same geometrical coverage as the galaxies. The simplest estimator involves counting pairs of galaxies as a function of separation and dividing over the same measure for random points [ref]:

$$\xi_g = \frac{\bar{n}_R}{\bar{n}_D}\frac{DD}{RR} - 1, \tag{1.42}$$

where DD and RR are the number of galaxy and random pairs, respectively, with a 3D separation between r and $r + \delta r$, and $\bar{n}_{D,R}$ are the galaxy and random number densities. The estimator can be improved, for instance to better handle edge effects, which can affect the clustering at large separations. The most common estimator nowadays is the LS estimator, which takes the form

$$\xi_g = \frac{1}{RR}\left[DD\left(\frac{\bar{n}_R}{\bar{n}_D}\right)^2 - 2DR\left(\frac{\bar{n}_R}{\bar{n}_D}\right) + RR\right], \tag{1.43}$$

which also uses the number of galaxy-random pairs at a given separation, DR. Due to the importance of random-random and galaxy-random pair counting in the equations above, it is of capital importance to construct a set of random points which accurately mimics the geometrical selection effects of the data. In addition, the random points should also follow the same redshift distribution as the galaxies. In order to minimize the shot noise from random catalogs, the standard approach is to create many of them, typically several times the number of galaxies

in hand.

As we will see in the next chapter, obtaining accurate 3D positions of galaxies is observationally difficult and expensive, as one has to rely on spectroscopy to obtain the redshifts of galaxies, providing their line-of-sight position. One common approach, also followed in this thesis, is to use photometry instead, which yields only approximate redshift information but at much faster rate. Therefore, we will not use the three-dimensional 2PCF, but rather the galaxy angular 2PCF in this work. In analogy to the 3D case, the galaxy angular 2PCF, $w(\theta)$, is defined as the probability above Poisson of finding two galaxies within the solid angle $d\Omega$ at an angular separation θ:

$$dP = \bar{N}(1 + w(\theta))d\Omega, \tag{1.44}$$

where $\bar{N}$ is now the mean number of galaxies per unit area. It is in practice estimated using again the Landy & Szalay estimator in Eq. (1.43), with galaxy and random pairs are now found with an angular separation between θ and $\theta + d\theta$. The angular 2PCF can be connected to the 3D power spectrum via the following relation:

$$w(\theta) = \int_0^\infty P(k)g(k\theta)kdk \tag{1.45}$$

with the kernel $g(k\theta)$ defined as:

$$g(k\theta) = \frac{1}{2\pi}\int_0^\infty dz J_0(k\theta\chi(z))\left(\frac{dN}{dz}\right)^2 \frac{dz}{d\chi}F(z) \tag{1.46}$$

(Limber 1953; Baugh & Efstathiou 1993). In the expression above, $\chi(z)$ is the comoving distance to redshift z, defined in Eq. (1.20), J_0 is the Bessel function of order zero and $F(z)$ is a function encapsulating the redshift evolution of density fluctuations in the given galaxy sample. Importantly, dN/dz is the redshift probability distribution in the galaxy sample, which contains the radial information needed in the 3D to angular 2D projection.

1.5.5 THE GALAXY BIAS

At this point, we have a way of comparing theory to observations through the clustering of galaxies. However, most of the cosmological information is contained in the matter correlation function (or the matter power spectrum), so we need to connect the two. Fortunately, as a consequence of the way structure forms in the Universe, the distributions of galaxies and dark matter are correlated: galaxies form within dark matter halos, which are the highest-density regions in the dark

matter distribution. Therefore, galaxies trace the matter distribution, but they are generally not unbiased tracers. This is usually expressed in terms of the so-called galaxy bias, b, which relates the galaxy and matter density fluctuations and can be generally written as

$$\delta_g = b\,\delta. \tag{1.47}$$

This relation translates into a connection between the galaxy and matter 2PCFs:

$$\xi_g(r) = b^2\,\xi_m(r), \tag{1.48}$$

which, in turn, directly translates to the corresponding angular 2PCF version:

$$w_g(\theta) = b^2\,w_m(\theta). \tag{1.49}$$

In general, the galaxy bias will be a function of scale (or angular separation in the 2D case) and redshift, so that, for instance, $b = b(\theta, z)$ in Eq. (1.49). For a given galaxy sample, with a specific redshift selection, the galaxy bias will be constant as a function of scale at large enough separation, where the density field follows linear evolution. In addition to the general galaxy bias, there may exist further differences between the galaxy and matter distributions, which are parametrized in terms of the so-called cross-correlation parameter, r, which may also depend on scale and redshift, and is defined as:

$$r = \frac{\xi_{gm}}{\sqrt{\xi_m\,\xi_g}}, \tag{1.50}$$

where ξ_{gm} is the galaxy-matter cross-correlation function, which is then related to the matter correlation function:

$$\xi_{gm} = b \cdot r\,\xi_m. \tag{1.51}$$

On large enough scales, in the linear regime, we expect the galaxy and matter distributions to trace the same structures and hence the cross-correlation coefficient to approach $r = 1$. In that case, the galaxy-matter 2PCF is simply $\xi_{gm} = b\,\xi_m$. Therefore, due to the different dependencies on the galaxy bias, having together the galaxy and galaxy-matter 2PCFs we are able to break degeneracies between the galaxy bias and the matter 2PCF, and hence access the cosmological information contained in the latter as well as information about how galaxies populate dark matter halos. In order to access the galaxy-matter 2PCF, galaxy clustering is not sufficient and we will need a different cosmological probe, the so-called weak gravitational lensing effect, which is the subject of the next section.

1.6 WEAK GRAVITATIONAL LENSING

General relativity describes how space-time curves under the influence of matter and energy. As particles (including photons) follow geodesics of space-time, light from distant sources will bend when passing around foreground mass-energy concentrations. Consequently, the images of distant galaxies are distorted due to the presence of foreground matter structures, an effect that is known as gravitational lensing. We can extract key information from gravitational lensing, as the distortions in the light are produced by the total matter gravitational field, not only by the matter we can see (e. g. galaxies or dust). Then, by quantitatively measuring these distorsions we can infer information about the mass distribution in the universe. The importance of that knowledge lies in the fact that, without lensing, all we know about the mass distribution of the universe comes from the galaxy distribution, and then it is subject to our knowledge about how light traces matter, i. e. our understanding of the galaxy bias described in the previous section. Using gravitational lensing, we can directly compare observations to theory avoiding the problem of galaxy bias, or we can learn about it to extract the most information from the combination of galaxy clustering and gravitational lensing observations.

Although gravitational lensing, in some extreme cases, can produce very impressive distortions, these cases are not very useful to obtain cosmological information out of them. Extracting large-scale structure information from gravitational lensing requires the study of faint distortions in the shapes of distant galaxies induced by foreground mass overdensities. We are not interested in any single overdensity (corresponding to, for instance, a cluster of galaxies) but rather in the large-scale properties of them throughout the universe. In this way, we do not need to infer exactly the mass distribution but we want to measure some simple statistics, as the correlation function and its Fourier transform, the power spectrum. Measuring these quantities is very useful since they enable us to directly compare cosmological theory, that predicts the behavior of these simple LSS statistics, with observations.

The typical gravitational lens system is depicted in the diagram of Fig. 1.7, with a mass concentration at redshift z_d (corresponding to an angular diameter distance D_d) deflecting the light rays from a source located at redshift z_s (at an angular diameter distance D_s). In this simple scenario, and provided the extent of the deflecting mass along the line-of-sight is much smaller than the distance to the lens plane, D_d, and the distance between the lens and source planes, D_{ds}, the light rays path, which is smoothly curving in reality, can be approximated as two straight rays with a sharp twist near the deflecting mass. The magnitude and direction of such twist are described by the so-called deflection angle, $\hat{\alpha}$, which is

determined by the deflecting mass distribution and the impact parameter ξ.

1.6.1 THE DEFLECTION ANGLE

In Sect. 1.1.2, we have introduced the FLRW metric, which provides a solution for the Einstein equations in a perfectly homogeneous universe. However, as gravitational lensing is produced by mass overdensities in the large-scale structure of the Universe, we need to consider light propagation in an inhomogeneous universe in order to study its effects. The line element for a general metric that describes an expanding Universe with first-order perturbations to its homogeneity is given by

$$ds^2 = \left(1 + \frac{2\Psi}{c^2}\right)c^2 dt^2 - a^2(t)\left(1 - \frac{2\Phi}{c^2}\right)dl^2, \tag{1.52}$$

where dl^2 is taken from Equations (1.2) and (1.4). The two potentials Ψ and Φ describe weak gravitational fields, with masses M and potentials $\Psi, \Phi \sim GM/R = (c^2/2)(R_s/R)$ whose extents R are much larger than their Schwarzschild radius R_s. In General Relativity and in the cases we will be considering, the two potentials are equal, $\Psi = \Phi$ (Kilbinger 2015). Also, if there are no perturbations, the metric reduces to the FLRW metric in Eq. (1.4).

Photons, as well as particles, will propagate through space-time following null geodesics, $ds^2 = 0$. Then, from the equation above and assuming weak gravitational fields we get the time of light ray travel to be:

$$t = \frac{1}{c}\int \left(1 - \frac{2\Phi}{c^2}\right)dr, \tag{1.53}$$

where the integral here is along the light path in physical coordinates r. From this equation, an analogy can be established between the gravitational potential and a medium with variable refractive index $n = 1 - 2\Phi/c^2$ (Schneider et al. 1992), and hence the name of gravitational lensing. Working with this optical lens analogy, we can use Fermat's principle, stating that light travels along the path that can be traversed in the least time, $\delta t = 0$. From this condition, and integrating along the light path we can obtain the expression for the deflection angle $\hat{\alpha}$, defined as the difference between the directions of the emitted and received light rays (see Fig. 1.7):

$$\hat{\alpha} = -\frac{2}{c^2}\int \nabla^p_\perp \Phi dr, \tag{1.54}$$

where the integration proceeds along the whole light path, in physical coordinates, and the gradient is taken perpendicular to it. Considering a point-like body

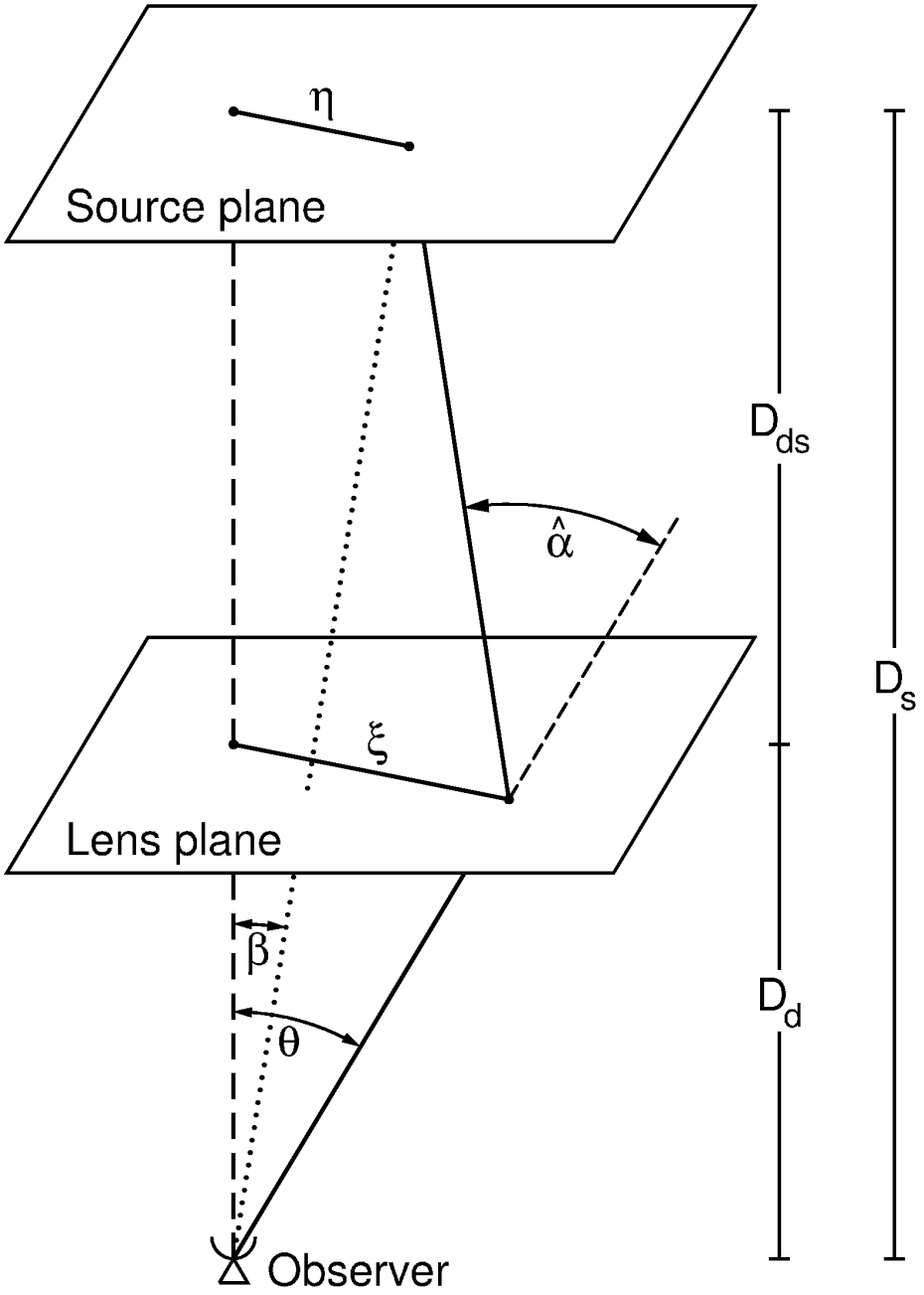

Figure 1.7: Diagram representing a typical gravitational lens system, showing the observer, lens and source planes, and (angular diameter) distances between them. From Bartelmann and Schneider (2001).

of mass M and its corresponding gravitational potential,

$$\Phi(\xi, z) = -\frac{GM}{\sqrt{\xi^2 + z^2}}, \tag{1.55}$$

where ξ is the impact parameter as described in Fig. 1.7 and z is the distance along the unperturbed light path, the deflection angles reduces to:

$$\hat{\alpha} = \frac{4}{c^2}\frac{GM}{\xi}. \tag{1.56}$$

This is usually expressed in terms of the so-called Schwarzchild radius, $R_S = 2GM/c^2$, simply as $\hat{\alpha} = 2R_S/\xi$. Importantly, this GR derivation using the metric in Eq. (1.52) yields a deflection angle which is a factor of two larger than the expected from purely Newtonian arguments. In 1919, during a solar eclipse, Arthur Eddington and collaborators measured the apparent change in position of stars as their light passed near the Sun, and proved the measurements to be consistent with the GR prediction (Dyson, Eddington & Davidson 1920).

1.6.2 THE LENS EQUATION

From the geometry of the gravitational lensing diagram in Fig. 1.7 we can relate the deflection angle to the other angles of interest, namely β, the angular position of the source galaxy had it not been lensed, and θ, the observed (lensed) position of the source galaxy, through the lens equation which also uses the relevant angular diameter distances:

$$\theta D_s = \beta D_s + \hat{\alpha} D_{ds}. \tag{1.57}$$

It is common to define the reduced deflection angle as

$$\alpha = \frac{D_{ds}}{D_s}\hat{\alpha}, \tag{1.58}$$

and then the lens equation simply becomes:

$$\beta = \theta - \alpha. \tag{1.59}$$

This equation relates the observed (lensed) position of a galaxy to its true (unlensed) position via the (reduced) deflection angle, and hence it plays a central role in the formulation of gravitational lensing. It can happen that, for fixed β, the equation has more than one solution, in which case the source galaxy has several images at different points in the sky. Such phenomenon is commonly referred

to as strong lensing and occurs when the so-called convergence field $\kappa(\theta) \geq 1$ somewhere, with the dimensionless convergence being defined as

$$\kappa(\theta) = \frac{\Sigma(\theta D_d)}{\Sigma_{\text{crit}}}. \tag{1.60}$$

Here $\Sigma(\theta D_d)$ is the surface mass density,

$$\Sigma(\theta D_d) = \int \rho(\theta D_d, z) dz, \tag{1.61}$$

where $\rho(\theta D_d, z)$ is the mass density and we have used $\theta D_d = \xi$ from Fig. 1.7, and

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}} \tag{1.62}$$

is the critical surface mass density, depending on both the lens and source redshifts, and sets the distinction between weak and strong lenses.

In terms of the convergence $\kappa(\theta)$, one can write the reduced deflection angle as

$$\alpha(\theta) = \frac{1}{\pi} \int d^2\theta' \kappa(\theta') \frac{\theta - \theta'}{|\theta - \theta'|^2}. \tag{1.63}$$

From this expression, it follows that the deflection angle can be expressed as the gradient of a potential,

$$\alpha(\theta) = \nabla \psi(\theta), \tag{1.64}$$

with

$$\psi(\theta) = \frac{1}{\pi} \int d^2\theta' \kappa(\theta') \ln \left| \theta - \theta' \right|^2 \tag{1.65}$$

being the deflection or lensing potential, the 2D analogue of the Newtonian potential, satisfying the Poisson equation as $\nabla^2 \psi(\theta) = 2\kappa(\theta)$.

1.6.3 DISTORTION AND MAGNIFICATION

From the lens equation we can obtain the different possible angular positions α of the images of a source at β. For an extended source, such as a typical galaxy, gravitational lensing can change its shape due to distinct light rays being deflected differently. When gravitational lensing is strong, for instance around massive galaxy clusters, the shape of source galaxies whose light passes near the cluster mass concentration can be distorted forming impressive and distinctive luminous arcs, as can be seen in Fig. 1.8. This is due to light rays coming from an extended source being more distorted the closer the pass to the cluster center (see Fig. 1.9

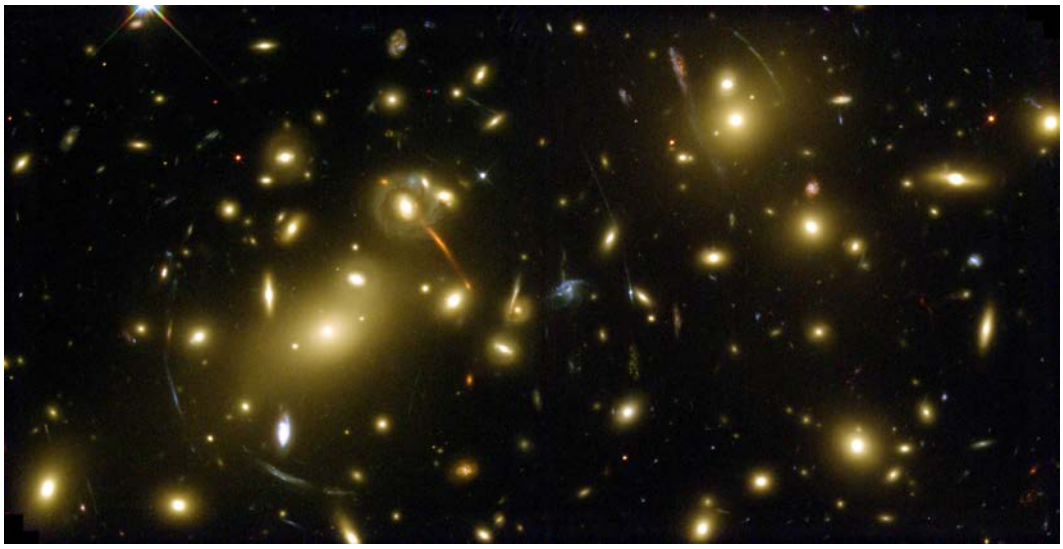

Figure 1.8: The massive foreground cluster (Abell 2218) distorting the images of background galaxies and forming arcs, due to strong gravitational lensing. The arcs are aligned in a way that their ellipticity is oriented tangent to the direction of the foreground mass, in this case the galaxy cluster center. Image credit: NASA/ESA.

for a graphical representation).

In general, the observed surface brightness distribution I is related to that at the source plane I^s by the following mapping:

$$I(\theta) = I^s[\beta(\theta)], \tag{1.66}$$

where surface brightness is conserved due to the absence of emission and absorption of photons in gravitational lensing. If the extent of a source is much smaller than the scale in which the lens properties vary, we can linearize the lens equation and then the distortion of the image can be described by the amplification matrix A, defined as

$$A = \frac{\partial \beta}{\partial \theta}. \tag{1.67}$$

Using the lens equation and the definition of the lensing potential,

$$A_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial}{\partial \theta_j}\left(\frac{\partial \psi}{\partial \theta_i}\right) \tag{1.68}$$

It is then convenient to parametrize the amplification matrix in terms of the scalar convergence κ and a two-component shear field $\gamma \equiv \gamma_1 + i\gamma_2 = |\gamma|\, e^{2i\phi}$ as:

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{1.69}$$

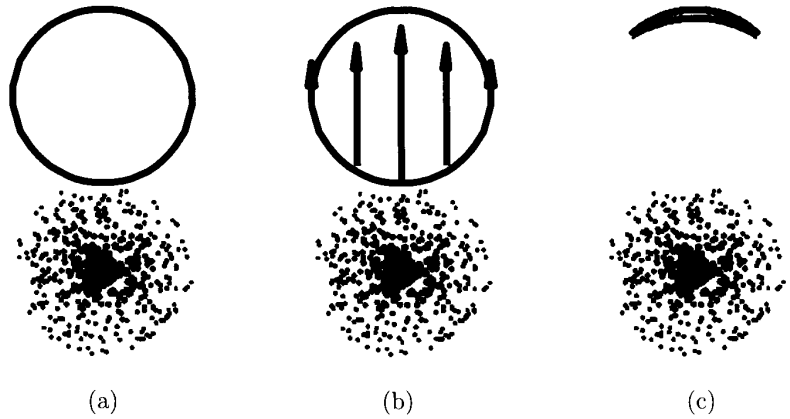

Figure 1.9: (a) Source circular galaxy placed behind a foreground mass distribution, represented as black points at the bottom. The observer is located where the reader is, so that the mass distribution lies between the observer and the source. (b) Light rays from the source get deflected as they pass near the mass distribution. The closer they pass to the center of the mass overdensity, the more deflected they get. (c) The resulting image is an arc, similar to the ones aparent in Fig. 1.8. From Dodelson (2003).

In this parametrization, both the shear and the convergence can be expressed as second derivatives of the lensing potential ψ:

$$\kappa = \frac{1}{2}(\partial_1\partial_1 + \partial_2\partial_2)\psi = \frac{1}{2}\nabla^2\psi$$
$$\gamma_1 = \frac{1}{2}(\partial_1\partial_1 - \partial_2\partial_2)\psi$$
$$\gamma_2 = \partial_1\partial_2\psi.$$

In order to understand the lensing effect on the observed source image, let's consider θ_0 to be a point within that observed image, corresponding to a point $\beta_0 = \beta(\theta_0)$ on the source. Then, using the linearized lens equation, the observed surface brightness (1.66) becomes:

$$I(\theta) = I^s[\beta_0 + A(\theta_0)\cdot(\theta - \theta_0)]. \tag{1.70}$$

Given this mapping, the images of circular sources become ellipses under the effect of gravitational lensing, and the properties of such ellipse are related to the matrix $A(\theta_0)$ (Bartlemann and Schneider 2001). On one hand, the ratios of the semi-axes of the ellipse to the radius of the source are given by the inverse of the eigenvalues of $A(\theta_0)$, $1-\kappa\pm|\gamma|$. On the other hand, the ratio between the solid angles subtended by the observed and the unlensed source images, called the

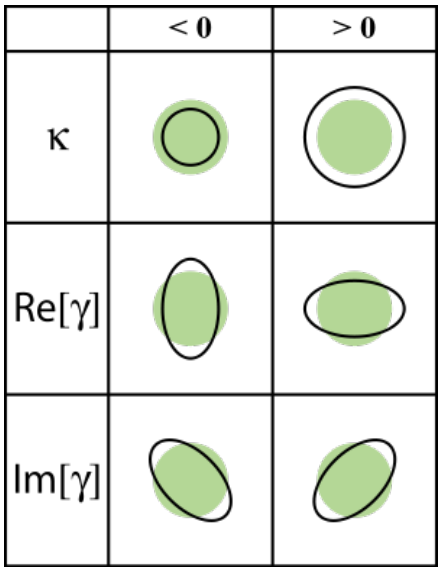

Figure 1.10: Effect of the convergence (κ) and the real (Re) and imaginary (Im) parts of the shear ($\gamma \equiv |\gamma| e^{2i\phi}$) on a circular source, represented in the diagram as a solid green circle. Image from Wikipedia.

magnification μ, is given by the inverse of the determinant of A:

$$\mu = \frac{1}{\det A} = \frac{1}{(1-\kappa)^2 - |\gamma|^2}. \tag{1.71}$$

Therefore, gravitational lensing distorts the source images in both shape and size (see Fig. 1.10). In addition, when estimating the shear of a galaxy, measurements are based in the shapes of galaxies, and do not consider their sizes. In that context, it is useful to rewrite A as

$$A = (1-\kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix}, \tag{1.72}$$

where we can see the $(1-\kappa)$ term only affects the size and not the shape of the observed image. Then, when using the shape of galaxies to estimate the shear, the observable is the reduced shear g and not the shear γ:

$$g = \frac{\gamma}{1-\kappa}. \tag{1.73}$$

In weak gravitational lensing, the values of the shear and the convergence are of the order of a few percent, $\gamma, \kappa \ll 1$, and, therefore, the reduced shear is a good approximation to the shear.

1.6.4 THE LENSING CONVERGENCE AS THE PROJECTED MATTER DENSITY

We have already presented the lensing potential as a 2D analogue of the 3D Newtonian potential. Using the 2D Laplacian and the 3D Poisson equation in

comoving coordinates, we can see the analogy as

$$\nabla^2\psi = 2\kappa \qquad , \qquad \nabla^2\Phi = \frac{3H_0^2\Omega}{2a}\delta. \tag{1.74}$$

Then, we can also think of the convergence κ as a 2D projected analogue of the matter overdensity δ. This relation, assuming the Universe to be flat for simplicity, can be expressed in the following way (Bartlemann & Schneider 2001):

$$\kappa(\theta,\chi) = \frac{3H_0^2\Omega_m}{2c^2}\int_0^{\chi} d\chi' \frac{\chi'(\chi-\chi')}{\chi}\frac{\delta(\chi'\theta,\chi')}{a(\chi')}. \tag{1.75}$$

For a redshift distribution of sources $p_z(z)dz = p_\chi(\chi)d\chi$, the convergence becomes:

$$\begin{aligned}\kappa(\theta) &= \int d\chi\, p_\chi(\chi)\,\kappa(\theta,\chi) = \\ &= \frac{3H_0^2\Omega_m}{2c^2}\int_0^{\chi_h} d\chi\, g(\chi)\,\chi\,\frac{\delta(\chi\,\theta,\chi)}{a(\chi)},\end{aligned} \tag{1.76}$$

where χ_h is the comoving horizon distance, defined as the comoving distance obtained for infinite redshift, and $g(\chi)$ is the source-redshift weighted lens efficiency factor:

$$g(\chi) = \int_{\chi}^{\chi_h} d\chi'\, p_\chi(\chi')\,\frac{(\chi'-\chi)}{\chi'}, \tag{1.77}$$

which indicates the lensing strength at a distance χ of the combined background galaxy distribution.

Chapter 2

THE DARK ENERGY SURVEY

We have seen in the last chapter that in order to constrain and unravel the nature of dark energy we need large data sets of galaxies with not only positions (in the sky plane) but also with redshifts measured. In this way, we intend to measure the large scale properties of the Universe by surveying large volumes and detecting the galaxies on them. The projects designed for such task are called galaxy redshift surveys and over the history of cosmology there are many examples of such observations. For instance, we can think of the Hubble's 1929 discovery of the expansion of the universe (§1.2.1) being the result of the first successful galaxy survey, and we have already mentioned the CfA galaxy redshift survey as providing the first evidence of the fact that galaxies lie on bubble-like structures with clusters of galaxies at their junctions. More recently, other galaxy surveys have proved successful in constraining cosmology, especially the 2-degree Field Galaxy Redshift Survey (2dFGRS[1]) and the Sloan Digital Sky Survey (SDSS[2]), and yet there are some ongoing and proposed projects expected to shed more light about the present unknowns in cosmology.

The Dark Energy Survey (DES) is an international, collaborative effort to study the origin of the accelerating expansion of the Universe and hence help understand the nature of dark energy by measuring the history of cosmic expansion as well as the history of the growth of structure in the Universe to high accuracy. More than 400 scientists from over 25 institutions in the United States, Spain, the United Kingdom, Brazil, Germany, Switzerland and Australia are working on the project (Fig. 2.1). This international collaboration has built and is using an extremely sensitive 570-Megapixel digital camera, DECam (Flaugher et al. 2015, Fig. 2.2), mounted at the prime focus of the 4-meter Blanco telescope at Cerro Tololo Inter-American Observatory (CTIO), in Chile.

[1] http://www.2dfgrs.net
[2] http://www.sdss.org

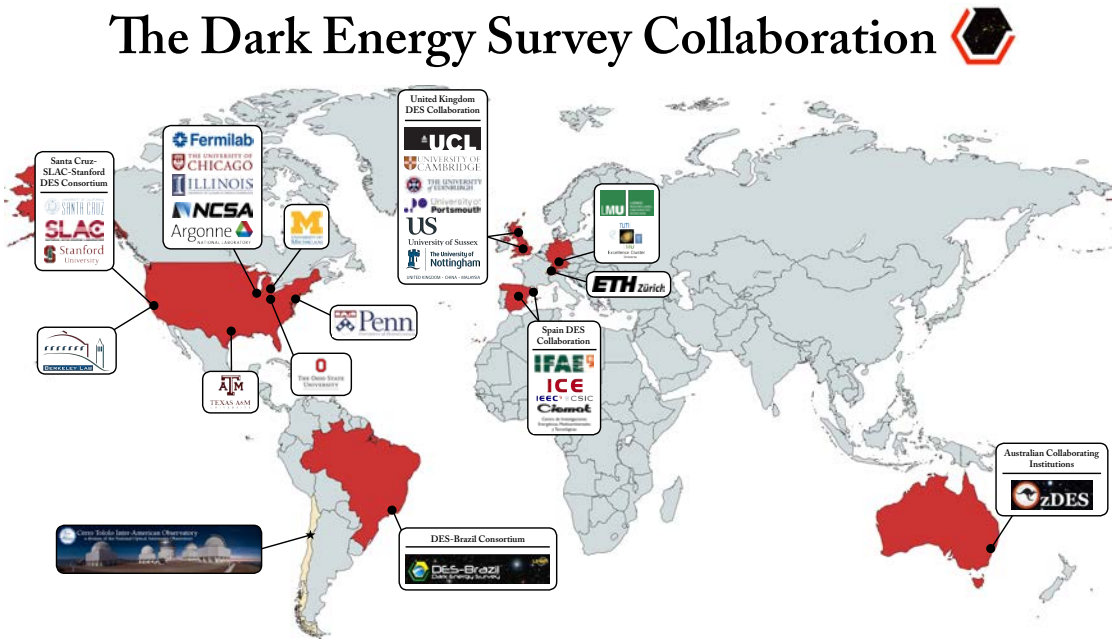

Figure 2.1: Map of the institutions involved in the Dark Energy Survey. Credit: Judit Prat.

The construction of the instrument started in 2008 and finished in 2011. The first assembly and testing were performed at Fermilab on an exact reproduction of the telescope structure, and then the camera and related components were shipped to CTIO between 2010 and 2011. DECam installation at the Blanco started in February 2012 and finished in September 2012, and it was led by CTIO and the DECam design and construction team. The DECam imager consists of a set of 74 state-of-the-art CCDs (charge-coupled devices), specifically designed by scientists at the Lawrence Berkeley National Laboratory (LBNL) to observe red light from distant galaxies, with thicknesses around 10 times larger than those of conventional CCDs to increase the chance of detecting long-wavelength photons. In addition, the camera has several other components:

- five optical lenses, one of which has a diameter of almost one meter and weighs 176 kilograms, making it the largest optical corrector component used in astronomy at the time of construction,
- five optical filters, spanning 400 nm to 1080 nm, also being the largest filters used in the astronomical community,
- a positioning and alignment system, consisting of an Hexapod mechanism with six pneumatically driven pistons which are used to align the optical lenses between exposures and to keep the camera on focus,

a) b)

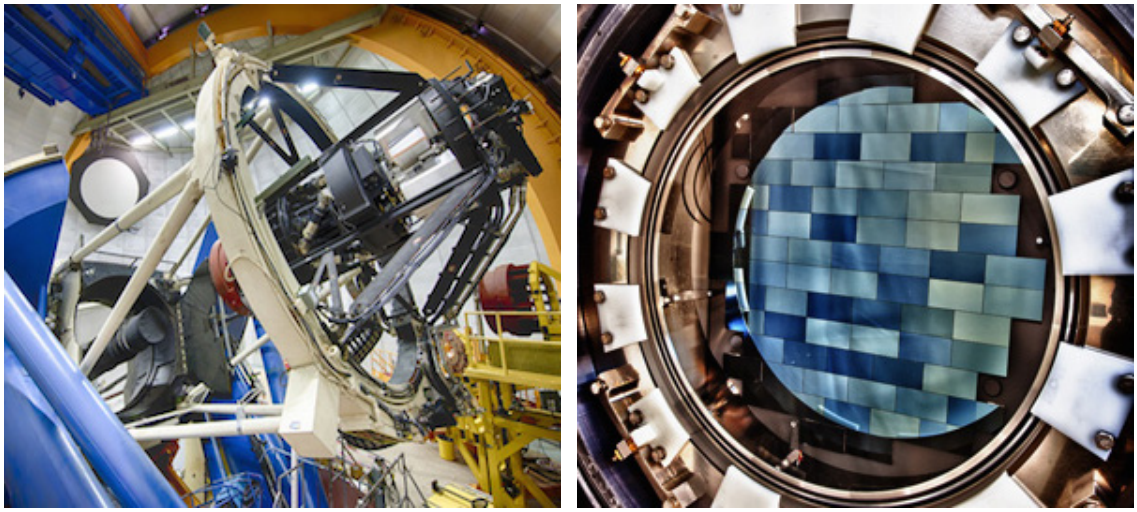

Figure 2.2: a) The Dark Energy Camera (DECam), mounted at the Blanco telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile. b) DECam features 62 charge-coupled devices (CCDs) for imaging and 12 CCDs for guiding and focus, which make a total of 570 megapixels (Flaugher et al. 2015). Credits: Reidar Hahn/Fermilab.

- a shutter, to shield the imager from light between exposures,
- and a cooling system using liquid nitrogen, which keeps the camera operating at -100^{o}C in order to minimize noise and dark current in the CCDs.

DECam saw its first light on the telescope on September 12, 2012. The commissioning of the camera took place during September and October 2012, and a Science Verification period of observations, covering around 200 deg.^{2} at the nominal depth of the survey lasted from November 2012 to February 2013. Starting August 2013, the DES Collaboration is using 525 nights of DECam observations over five years (2013-2018) to carry out a multi-band ($grizY$) imaging survey that will cover around 5000 deg.^{2} of the southern sky to a depth of around $i_{AB} = 24$ magnitude. By the end of the survey, DES will have measured positions, photometric redshifts (§2.1) for around 300 million galaxies and shapes (§2.2) for over 200 million of those galaxies. In addition, DES will detect over 100000 galaxy clusters and, by devoting a fractio of its time to observe smaller patches of sky roughly once a week, about 3000 type Ia SNe. Compared to previous galaxy surveys, DES represents an increase in volume over the Sloan Digital Sky Survey (SDSS) by roughly a factor of 7 (Flaugher et al. 2015). In this chapter we review two crucial techniques in a photometric survey like DES: the procedures to measure photometric redshifts and galaxy shapes.

2.1 SPECTROSCOPIC AND PHOTOMETRIC REDSHIFT SURVEYS

In galaxy redshift surveys, angular positions are measured directly through the positions of galaxies in the sky plane, while radial information in the line of sight is related to the redshift of individual objects. In this way, measuring angular positions for galaxies is straighforward, but obtaining radial information is not.

In Astronomy, the redshift of a distant celestial object is measured through the shift in its spectrum $S(\lambda)$. We call spectrum to the flux distribution in wavelength space, also known as spectral density flux (units of [energy]/([area] · [time] · [wavelength]). We compute the redshift by comparing a spectrum in the rest frame $S(\lambda)$ to a spectrum of a distant object that appears to be moving with respect to us due to the expansion of the universe $S'(\lambda';z)$. The transformation from the rest frame to a distant one from which we extract the redshift goes as follows:

$$S(\lambda) \to S'(\lambda';z) = \frac{S(\lambda = \lambda'/(1+z))}{(1+z)^2}, \tag{2.1}$$

where the wavelength transformation comes from equation (1.10). The factor $(1+z)^{-2}$ comes from two different facts. On the one hand, using again the wavelength transformation $\lambda' = (1+z)\lambda$ from equation (1.10) and the energy of a photon $E_\gamma \propto \lambda^{-1}$ we obtain $E' = E/(1+z)$. Since $S(\lambda)$ measures energy, we get the first $(1+z)^{-1}$ factor from here. On the other hand, the time between two consective photon receptions τ increases for an emitter object moving along the line of sight such that $\tau' = (1+z)\tau$. So, since $S(\lambda)$ has units of time^{-1} we get the second $(1+z)^{-1}$ factor from there. A graphical representation of equation (2.1) is shown in Figure 2.3, where we can see how spectra get decreased, stretched and moved to higher wavelenghts when redshift increases. However, the most commonly used detectors in astronomical observations are Charged Couple Devices (CCDs) and they are not sensitive to the energy of the detected photon, but only to the number of photon counts. Therefore, if we define the spectral density flux in number of counts units instead of energy, the transformation (2.1) will have only one $(1+z)^{-1}$ factor.

There are two main kinds of galaxy redshift surveys, corresponding to the way in which they observe galaxies and then related to the way in which they measure redshifts. Although in the past redshifts were mainly measured spectroscopically, recently there are many projects that are only based on imaging, using photometry to determine the redshift of each galaxy. On the one hand, the traditional spectroscopic redshift technique uses a spectrograph to measure redshifts with high accuracy but has some drawbacks like the amount of observing time required for those accurate determinations, or the need for previous galaxy targetting, with the

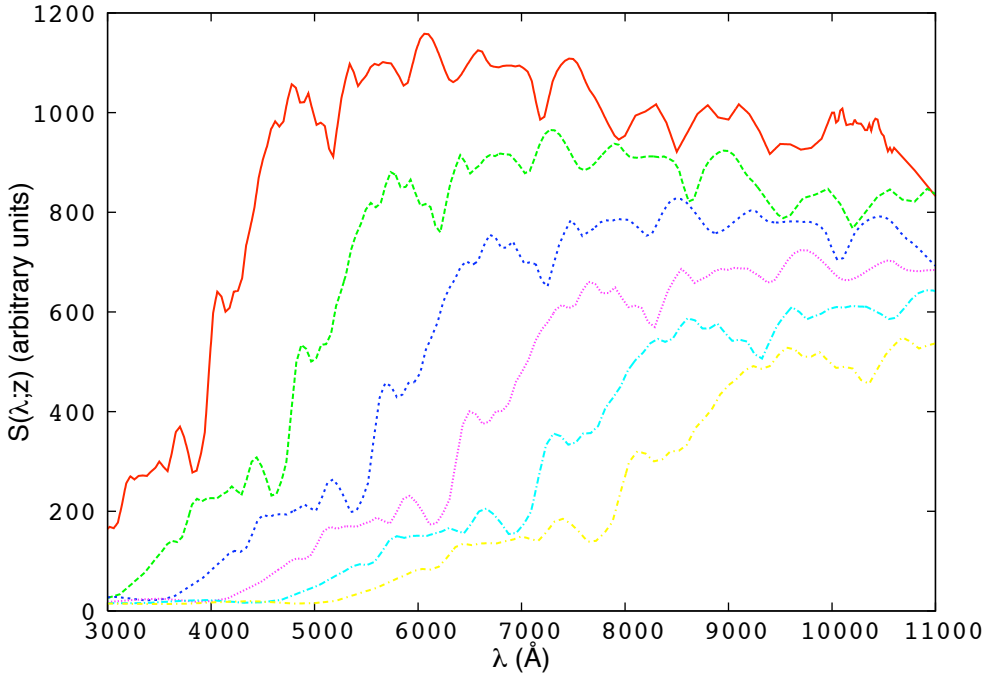

Figure 2.3: The different curves corresponding to the same spectrum at different redshifts. $z = 0.0$ for the solid red curve, $z = 0.2$ for the dashed green curve, $z = 0.4$ for the dashed blue curve and $z = 0.6,\ 0.8$ & 1.0 for the consecutive shifted curves.

risk of producing a biased sample. On the other hand, the photometric redshift technique is not generally comparable to the spectroscopic equivalent in accuracy in determining redshifts, but has the advantatge that you can obtain redshifts much faster in a systematic way, without the need of targetting.

As we have already sketched, the accuracy on measuring redshifts for individual galaxies is directly related to the radial information we get for them. In this way, we can consider that spectroscopic surveys provide 3D cosmological information while photometric surveys provide 2.5D information, as they also get some radial information, but not so precise.

2.1.1 SPECTROSCOPIC REDSHIFT DETERMINATION

With this technique, one starts the determination of the redshift of a given object from the measurement of its spectrum $S(\lambda; z)$ (e.g. one of the curves shown in Figure 2.3). The light from the object is separated into narrow wavelength bins a few Å across by dispersing the light with, for instance, a prism. Each bin recieves then only a small fraction of the total light from the object. Hence, to achieve a sufficiently high signal-to-noise ratio in each bin, long integration times are required.

Suppose we have measured the green curve in that figure and we want to estimate the redshift of that object taking into account the spectrum for the same class of galaxies, called in this example *Luminous Red Galaxies* (LRGs) and corresponding to the red curve in the same figure. We observe in that red curve a characteristic feature at 4000 Å, which consists on a abrupt increase of the spectral energy density. This feature is known as the 4000 Å break in LRGs. Now, we observe that the same feature in the green curve is observed instead at 4800 Å, so by equation (1.10) we compute the redshift for that object as:

$$z = \frac{4800 - 4000}{4000} = 0.2. \tag{2.2}$$

So, in general, if we are able to measure the spectrum of an object and also have information about the spectrum at rest for that given class of objects we can look at some features (the 4000 Å break in the example of LRGs but there are other examples like *Emission line galaxies*, ELGs, where you look at the positions of some narrow peaked lines) of the spectrum and by comparing the wavelengths at which we encounter those features we can measure the redshift for that object by applying the same simple equation (1.10), as already done in (2.2).

2.1.2 PHOTOMETRIC REDSHIFT DETERMINATION

The photometric redshift determination technique is a method to measure the redshifts of objects using imaging instead of spectroscopy. In this method, the photometry of observed objects is converted into low resolution spectra and, then, redshifts are determined by comparing these measured spectra to redshifted template galaxy spectra for that given class of objects or to some other low-resolution measured spectra for which we already have spectroscopy. The different photometric redshift methods used by the Dark Energy Survey, and the precision acquired with them using Science Verification data, are the subject of Part II of this thesis.

The main advantage of the photometric redshift technique is speed. In this case, we don't disperse light but we use filters[3], which effectively integrate all the light in a given wavelength range. These filters are usually wide (typically of the order of 100 nm). If each different filter i has throughput $R_i(\lambda)$ (the response of a filter as a function of λ with units of $distance^{-1}$), where i =1, ..., N = number of filters, then, instead of the spectrum, we will have the fluxes of the object in each

[3]In astronomy, a *filter* is some mechanism that allows the selection of only a given range of wavelengths for the electromagnetic spectrum of an object.

filter, being proportional to the following integral:

$$F_i \propto \int_0^\infty S(\lambda; z) R_i(\lambda) \lambda d\lambda. \tag{2.3}$$

The measured fluxes for all filters (up to normalization) are all we need to extract the redshift of the measured object, as we will see below. It is also important to stress that in astronomy usually one works with magnitudes instead of fluxes or counts. A magnitude is a logarithmic measure of the brightness of an object, and can also be measured in a specific filter[4]. Since we know the area of the telescope and the exposure times for each filter we may easily change from fluxes to magnitudes later on.

Since filters are usually wide enough, it requires only a short exposure time to reach a high signal-to-noise ratio measurement. Furthermore, imaging detectors usually cover a larger area of the sky compared to multi-object spectrographs, and hence they can measure redshifts of more objects at the same time. Not only that, but using imaging we obtain low-resoltion spectra for every galaxy detected in a given image, without the need of targetting a priori.

On the other hand, the main disadvantage of photometric redshifts is the low precision acquired in the redshift estimation. While spectroscopic redshifts have precision better than $\sigma_z = 0.001$, photometric redshifts usually have precision of the order of $\sigma_z = 0.1$.

There are two main different techniques for computing photometric redshifts. Both techniques start from having a low resolution spectra but they differ in the way to measure the redshift: either comparing it with redshifted spectral templates or comparing it with real galaxy low-resolution (photometric) spectra for which the redshift is known through additional spectroscopic measurements.

TEMPLATE-BASED TECHNIQUES

This technique uses a predefined set of theoretical spectra for different types of objects that are called *templates*. Then, the idea is to integrate these templates using the same integral as in equation (2.3) while redshifting the templates using (2.1), obtaining a set of fluxes for each band depending on redshift $\{F_i^{(tem)}(z)\}$. After that, we compute the redshift of the object by comparing the measured fluxes $\{F_i\}$ with the ones computed out of the redshifted templates $\{F_i^{(tem)}(z)\}$ and looking for the z that makes the match better. Taking into account that measured fluxes have associated an uncertainty as well, $\{F_i, \sigma_{F_i}\}$, we minimize the following χ^2

[4]The magnitude of an object is defined as $m = -2.5 \log_{10} \frac{F}{F_0}$, where F is the measured flux of the object and F_0 is a flux of reference.

function in order to obtain the redshift of the object:

$$\chi^2(z) \equiv \sum_{i=1}^{N} \left[\frac{F_i - kF_i^{(tem)}(z)}{\sigma_{F_i}} \right], \tag{2.4}$$

where k is a normalization constant and N is the number of filters, typically around 4-6, although there are galaxy surveys using many narrower filters, as it is the case of the PAU (Physics of the Accelerating Universe) survey[5].

Here we have introduced the basics of the template-fitting method. Nevertheless, there are many ways to go beyond these basics in many aspects. For instance, one usual way to improve the redshift precision is to perform interpolation between many different templates for a given object. However, more important is the use of bayesian statistics to improve the quality of the fit by introducing some prior knowledge of the distribution of objects one expects to find (Benitez et al. 2000). In this framework one ends up having a probability density function (pdf) for the redshift of each galaxy, delivering more information about each single object, together with other advantages as, for example, recovering the whole redshift distribution for the whole sample.

Moreover, template-based methods are necessary in order to estimate redshifts for objects that are not accessible to spectroscopic measurements because they are too faint. While this technique to estimate redshifts is still necessary, it has some drawbacks like the need for a precise understanding of the relative efficiency of each filer band or the need for galaxy templates spanning the whole wavelength range and all spectral types involved.

TRAINING-BASED TECHNIQUES

The other widely used approach for computing photometric redshifts is more empirical. In that alternative case, we start by having a set of objects for which we know both spectroscopy and photometry that is known as *training sample* or *training set*. Then, for those specific objects we have measured fluxes in all bands but we also know their spectroscopic (i.e. high precision) redshifts. So we know $\{F_i, \sigma_i, z_i\}$ for the training sample, or $\{m_i, \sigma_{m_i}, z_i\}$ if we work in magnitude space. In fact, one usually works with *colors* instead of magnitudes, which are defined as differences between magnitudes in different bands. In an ideal case, this training set will span both the whole redshift and color space of our survey and it will be representative of the sample of objects we aim to measure. The training set could also be derived from a set of templates or from simulated catalogues.

[5]http://www.pausurvey.org

Given a representative training sample, the method consists in finding a transformation from points in the multidimensional color space to points in redshift space, i.e. a mapping function from color to redshift space. As soon as we have determined this transformation we will be able to apply such transformation for all the other measured objects for which we don't have spectroscopy (known sometimes as *testing sample* or *testing set*, those being the bulk of the survey objects) and obtain a redshift for all of them.

Setting up the mapping between the space of observed magnitudes and redshift space can be considered as a machine-learning problem. Many different methods can be utilized to find such mapping. The most popular examples are artificial neural networks, k-nearest neighbor algorithms, random forests or boosted decision trees.

Training-based methods are normally used when it is difficult to find representative templates for the objects, the parameter space of the observables is not too large and a large training sample is available. Moreover, some other advantages of those methods lie on the fact that they can also use other parameters (not only fluxes or magnitudes in each filter) to estimate the redshift of an object. For instance, they can use galaxy shapes or information about the environment as inputs in the estimation.

2.2 MEASURING GALAXY SHAPES

Similar to the need of measuring galaxy redshifts in order to access LSS information and probe cosmology using galaxy surveys, in order to study the weak gravitational lensing (WL) effect described in the previous chapter we need to measure the shapes of galaxies to excellent accuracy. That is another major advantage of photometric surveys over spectroscopic surveys: those of the latter type cannot measure galaxy shapes as, for that, images are needed.

We have seen that the WL effect produces distortions in the size and the shape of background galaxies, due to the gravitational field of foreground mass structures. We are most interested in the elliptical shape distortions produced by the shear field, as depicted in Fig. 1.10. In order to detect these, we need to precisely measure the ellipticity of galaxies detected in the survey images. For that purpose, we start by defining image ellipticities. Let $I(\vec{\theta})$ be the brightness distribution of an isolated galaxy image on the sky. Then, for instance, the center of the image is defined as

$$\bar{\vec{\theta}} \equiv \frac{\int d^2\theta\, I(\vec{\theta})\, q_I[I(\vec{\theta})]\, \vec{\theta}}{\int d^2\theta\, I(\vec{\theta})\, q_I[I(\vec{\theta})]}, \tag{2.5}$$

where $q_I(I)$ is a chosen weight function. Galaxy shapes are usually defined in

terms of the weighted second moments of the brightness distribution,

$$Q_{ij} = \frac{\int d^2\theta\, I(\vec{\theta})\, q_I[I(\vec{\theta})](\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int d^2\theta\, I(\vec{\theta})\, q_I[I(\vec{\theta})]}, \qquad i,j \in \{1,2\}. \tag{2.6}$$

For an image with circular isophotes, $Q_{11} = Q_{22}$, and $Q_{12} = 0$. The trace of Q relates to the size of the image, while the traceless part contains the ellipticity information. Then, one can define a complex ellipticity (Schneider 2006),

$$\epsilon \equiv \frac{Q_{11} - Q_{22} + 2\mathrm{i}Q_{12}}{Q_{11} + Q_{22} + 2\sqrt{Q_{11}Q_{22} - Q_{12}^2}}. \tag{2.7}$$

When the image has elliptical isophotes of minor-to-major ellipse axis ratio of $r \leq 1$, one obtains:

$$\epsilon = \frac{1-r}{1+r} e^{2i\phi}, \tag{2.8}$$

where ϕ is the position angle. Before gravitational lensing occurs, each galaxy has an intrinsic complex ellipticity ϵ^s, defined in analogy to (2.7). Then, ϵ^s is modified by the (reduced) gravitational shear g, yielding the observed ellipticity ϵ. In this way, the observed ellipticity as a function of the intrinsic ellipticity and the reduced shear (for $|g| \leq 1$) is (Schneider 2006)

$$\epsilon = \frac{\epsilon^s + g}{1 + g^* e^s}, \tag{2.9}$$

where the asterisk denotes complex conjugation. In the weak lensing regime, this expression can be approximated to

$$\epsilon \approx \epsilon^s + \gamma. \tag{2.10}$$

If the intrinsic ellipticities of the galaxies are randomly oriented, the mean of the observed ellipticity is an unbiased estimator of the shear, since $\langle \epsilon^s \rangle = 0$, and then

$$\langle \epsilon \rangle \approx \gamma. \tag{2.11}$$

In practise, this estimator is biased by the presence of intrinsic galaxy alignments, an effect given by the correlation of galaxy shapes in the absence of gravitational lensing, due solely to gravitational interactions between galaxies. Because galaxy ellipticities are no longer randomly oriented, intrinsic alignments can add an excess of correlation between galaxy shapes and introduce a bias in gravitational lensing analyses.

In the weak lensing regime, the shear of a galaxy cannot be estimated on a

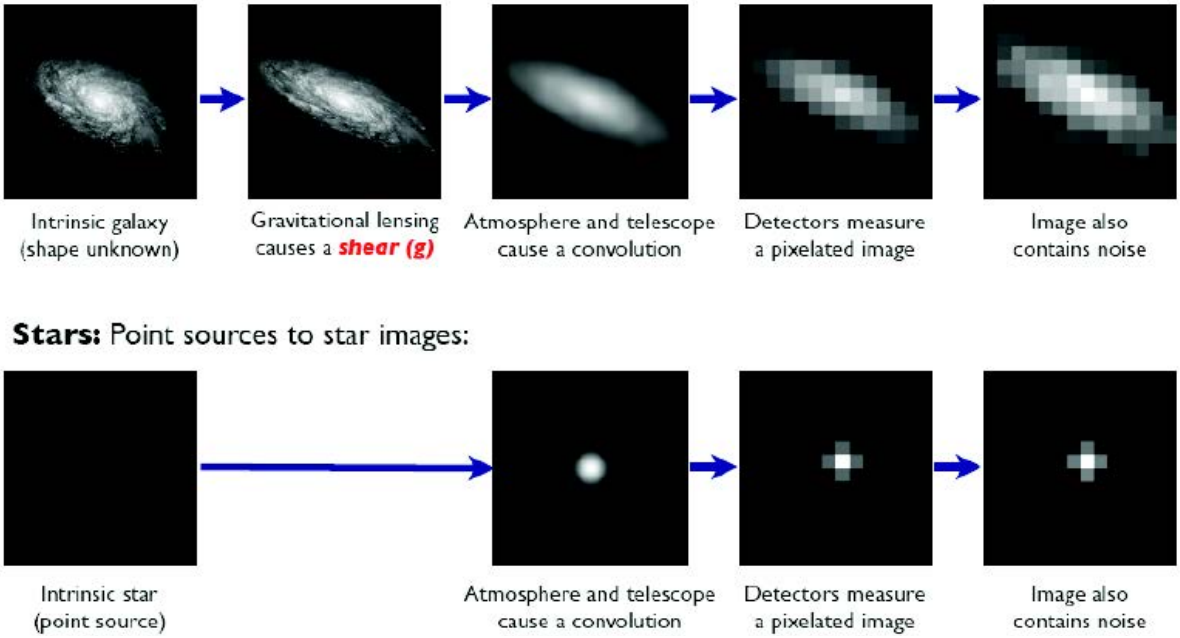

Figure 2.4: Forward process in shear measurements. The upper row is showing the different effects occuring to a galaxy from its intrinsic shape to the observed image. The lower row shows the same effects occuring on stars (note there is no gravitational lensing shear for stars). The main task in shape estimation is to recover an unbiased estimate of the shear of galaxies using the information in the last image of each row. Credit: Bridle et al. (2009).

galaxy-by-galaxy basis. This is because the typical distortions produced by LSS induced gravitational lensing are of the order of $\gamma \sim 0.03$, much smaller than the typical intrinsic ellipticity scatter $\sigma_\epsilon^s = \left\langle |\epsilon|^2 \right\rangle^{1/2} \sim 0.3$ (Kilbinger 2014).

Observationally, gravitational lensing is not the only effect altering the intrinsic shape of galaxies when they are detected. Effectively, galaxy images are also convolved by a kernel, known as the Point Spread Function (PSF), which describes the response of the telescope to a point-like source. An isotropic PSF will make galaxies look rounder, hence erasing the shear effect, while anisotropies in the PSF will directly contaminate the shear signal by making galaxies look more elongated in the direction of the anisotropy. Figure 2.4 shows a visualization of the different effects altering the intrinsic image of a galaxy when observed, from gravitational lensing to PSF convolution, and even pixelization in the imager and the corresponding noise.

In this context, in order to measure the shapes of galaxies to perform a weak gravitational lensing measurement we need to deconvolve the effect of the PSF on the galaxy images to recover an unbiased estimate of the shear. For this purpose, a model for the PSF along the camera field of view is created for every exposure. That is performed by identifying a number of stars in each of these exposures and characterizing their detected shape, as sketched in the lower row of Figure 2.4.

An accurate model for the PSF and its variations along the field of view is crucial to correct for circularizing effects as well as biases coming from potential PSF anisotropies. Optical effects in the telescope should be properly included in the modeling as well. A correct deconvolution of the PSF is another crucial step, as the PSF ellipticity can be as large as 10% while the shear effect can be an order of magnitude smaller. Furthermore, the PSF needs to be interpolated to the different galaxy locations using the limited number of stars available in each exposure.

In DES, the survey area, the observed depth and the size of the PSF will allow the Collaboration to measure over 200 million galaxy shapes, with an effective number density of around 10 galaxies per arcmin^2. That will be the largest data set ever used in any weak gravitational lensing analysis, bringing down statistical errors to an unprecendented level, and making it crucial to control systematic errors so that they do not dominate the error budget.

Part II

Photometric redshifts

Chapter 3

PHOTOMETRIC REDSHIFTS IN THE DES-SV DATA

3.1 INTRODUCTION

Large galaxy surveys provide detailed information on the large-scale structure of the Universe, which, in turn, helps understand its geometry, composition, evolution and fate. On one hand, spectroscopic surveys like 2dF (Colless et al., 2001), VVDS (Le Fèvre et al., 2005), WiggleZ (Drinkwater et al., 2010) or BOSS (Dawson et al., 2013) provide a three-dimensional picture of the galaxy distribution, but they are costly in time and resources, and may suffer from limited depth, incompleteness and selection effects. On the other hand, photometric surveys such as SDSS (York et al., 2000), PanSTARRS (Kaiser, Tonry & Luppino, 2000), KiDS (de Jong et al., 2013), HSC[1] or LSST (Tyson et al., 2003) are more efficient and usually deeper, more complete and nearly unbiased, but do not provide a complete 3D view of the Universe, due to their limited resolution in the galaxy positions along the line of sight, which are computed by measuring the photometric redshift (photo-z) of each galaxy from the fluxes measured through a set of broadband filters. Even with their limited resolution along the line of sight, photometric surveys, because of their larger volume, are extremely useful for cosmology and, furthermore, uniquely provide some of the most stringent probes of dark energy, such as weak lensing.

There are two main approaches for measuring photometric redshifts: template fitting methods (e.g. `Hyperz`, Bolzonella, Miralles & Pell (2000); `BPZ`, Benitez (2000); Coe et al. (2006); `LePhare`, Arnouts et al. (2002); Ilbert et al. (2006); `EAZY`, Brammer, van Dokkum & Coppi (2008)), in which the measured broadband galaxy spectral energy distribution (SED) obtained from the fluxes is compared to a set of redshifted galaxy templates until a best match is found, thereby determining both the galaxy spectral type and its redshift; and training methods (e.g.

[1] http://www.naoj.org/Projects/HSC/index.html

`ANNz`, Collister & Lahav (2004); `ArborZ`, Gerdes et al. (2010); `TPZ`, Carrasco Kind & Brunner (2013)), in which a set of galaxies with known spectroscopic redshifts is used to train a machine-learning algorithm (an artifitial neural network, for example), which is then applied over the galaxy set of interest. Each technique has its own advantages and disadvantages, as we will discuss in this chapter, and a combination of them can fully exploit this fact (Carrasco Kind & Brunner, 2014).

In order for photo-z's to be useful for cosmological studies, it is necessary to calibrate them, by understanding the statistical properties of the distribution of the differences between the true galaxy redshifts and their photo-z estimates: its mean value (for the bias), its width (for the resolution), and its tails (for the fraction of outliers, with grossly misestimated photo-z's). To accomplish this, a sample of galaxies with spectroscopic redshifts is required, ideally with a galaxy population that reproduces the population in the photometric survey.

The Dark Energy Survey (DES, Flaugher (2005)) is one such photometric redshift survey, and will cover about one eighth of the sky (5000 sq. deg.) to an unprecedented depth ($i_{AB} < 24$), imaging about 300 million galaxies in 5 broadband filters ($grizY$) up to redshift $z = 1.4$. The DES camera (DECam, Flaugher et al. (2012); Diehl et al. (2012)) was installed and commissioned in the second semester of 2012, and a Science Verification (SV) period of observations followed, lasting from November 2012 to February 2013. The survey officially started in late August 2013.

The SV observations provided science-quality data for almost 200 sq. deg. at close to the nominal depth of the survey. The SV footprint was chosen to contain areas already covered by several deep spectroscopic galaxy surveys, including VVDS (Le Fèvre et al. (2005)), ACES (Cooper et al. (2012)), and zCOSMOS (Lilly et al. (2007)), which together provide a suitable calibration sample for the DES photometric redshifts. This chapter presents a study of the photo-z precision achieved by DES during the SV period, by taking advantage of the available spectroscopic data in its footprint, and by using a large number of photo-z algorithms of different nature.

It has been pointed out (Cunha et al., 2012a) that cosmic variance in the spectroscopic samples used for photo-z calibration may bias the results of an analysis such as the one we present here, which uses spectra in four relatively small (1 sq. deg. each) patches of sky. A robust photo-z calibration requires galaxy spectra distributed all over the photometric survey's footprint, calling for as many as O(50–100) patches (Cunha et al., 2012a) . While the plan for the ultimate photo-z calibration of the whole DES data will need such a spectroscopic calibration sample, and steps are being taken towards the acquisition of the relevant data, the currently available spectroscopic data set is good enough for a first analysis of the photo-z precision that can be achieved with the early DES data. Analogously,

the ultimate DES photo-z calibration will have to worry about the effects of the possible incompleteness of the spectroscopic calibration samples, effects that we can safely ignore here, given the scope of this first study.

Many studies have been performed in the past comparing in detail several photo-z codes over the same, real or simulated, data set (Hogg et al., 1998; Abdalla et al., 2011; Hildebrandt, Wolf & Benítez, 2008; Hildebrandt et al., 2010; Dahlen et al., 2013). Particularly comprehensive is the work by Hildebrandt et al. (2010), which compares the performance of 19 photo-z codes both over simulated and real (including HST) observations taken in 18 optical and near-infrared bands. Similarly, Dahlen et al. (2013) compares 11 codes over real data in 14 bands, including also some HST data. On the other hand, Abdalla et al. (2011), analyzed the performance of six photo-z algorithms on the MegaZ Luminous Red Galaxy sample extracted from the SDSS Data Release 7 five-band photometry, with a magnitude limit around $i_{AB} = 20$. This chapter differs from these previous studies in that, on the one hand, it uses solely DECam five-band photometry ($grizY$), and on the other, it studies all kinds of galaxies up to the DES nominal limiting magnitude $i_{AB} = 24$. Furthermore, in the present study, rather than trying to carry out a thorough comparison of all the photo-z codes available in the literature, we concentrate on assessing the performance of the early DES data with respect to the photometric redshift determination, and, in order to do so, we try the codes in which members of the DES collaboration have a certain degree of expertise, without attempting to be complete or even necessarily fair in the comparison. Beyond providing a snapshot of the quality of the DES-SV data regarding photo-z estimation and accuracy, a secondary goal of this work is to tune these photo-z codes to the particular characteristics of the DES data: filter set, depth, etc, in preparation for the upcoming larger data sets.

Since even the deep spectroscopic samples mentioned above fail to reproduce exactly the depth and colors of the DES-SV photometric galaxy sample, a multi-dimensional weighting technique (Lima et al. (2008); Cunha et al. (2009)) was used in order to bring the spectroscopic and photometric samples into better agreement. Matching the galaxies in the spectroscopic samples with those in the DES-SV photometric sample and comparing their spectroscopic redshifts with the DES photo-z's, we will show that, even at this early stage, the DES-SV data fulfill the set of photo-z requirements on bias, resolution and outlier fraction that were defined prior to the start of the survey.

The outline of the chapter is as follows. Section 2 describes the DES-SV photometric galaxy sample, whereas the spectroscopic galaxy samples are presented in Section 3, together with the weighting technique that has been used to match their depth and colors to those of the DES-SV sample. Section 4 describes briefly the conditions in which the 13 different photo-z codes studied were run, and con-

tains the bulk of the results of the chapter, including the comparison between the results obtained with the different photo-z codes, the dependence of the results on both the depth of the DES-SV data and the specific spectroscopic calibration samples used, and an in-depth presentation of the results obtained with four representative photo-z codes, in particular with respect to the set of requirements of the DES survey, which we set up at the beginning of Section 4. A discussion of the main results in the chapter can be found in Section 5. Finally, we present our conclusions in Section 6, while we confine to an appendix the detailed description of the metrics used to characterize the photo-z distributions.

3.2 DES-SV PHOTOMETRIC SAMPLE

DECam imaging on fields overlapping those from deep spectroscopic redshift surveys were obtained for the following four DES fields: SN-X3, SN-C3, VVDS F14, and COSMOS, whose positions in the sky are shown in Fig. 3.1. SN-X3 and SN-C3 are the two deep fields in the DES supernova survey, and dithered observations of these fields were obtained routinely during the DES SV period. The SN-X3 field includes the VVDS-02hr field of the VVDS Deep survey (Le Fèvre et al., 2005, 2013), while SN-C3 overlaps with the CDFS (Chandra Deep Field South) area of the ACES survey (Cooper et al., 2012). The VVDS F14 field was centered on the VVDS-Wide redshift survey 14hr field (Garilli et al., 2008), and dithered imaging to DES main survey depth of this field was likewise obtained during DES SV. Deep dithered imaging data for the COSMOS field, centered on the Cosmological Evolution Survey (COSMOS) area (Lilly et al., 2007, 2009) were obtained during February 2013 by a DECam community program.[2] Each one of the four fields covers about the area of a single DECam pointing, or about 3 deg^2. See Section 3.3 for a detailed description of the spectroscopic data matched in each of the fields.

All fields include imaging in the 5 DES filters $grizY$, and additionally in the u band, which is part of DECam but not used by the DES survey. The data have been processed to two imaging depths: Main, corresponding to approximately DES main survey exposure times, and Deep, corresponding to about 3 times the exposure of a single visit to a DES supernova deep field (for SN-X3 and SN-C3) or deeper (for COSMOS). Differences in S/N between the Main and Deep samples can be appreciated in Fig 3.2; details of the data, the exposure times used and the magnitude depths are given in Table 3.1. Similar to DES science requirements convention, the 10σ magnitude limit is defined to be the `MAGAUTO` value (see

[2]Proposal 2013A-0351 Made available for DES photo-z calibration use by arrangement with PI Arjun Dey.

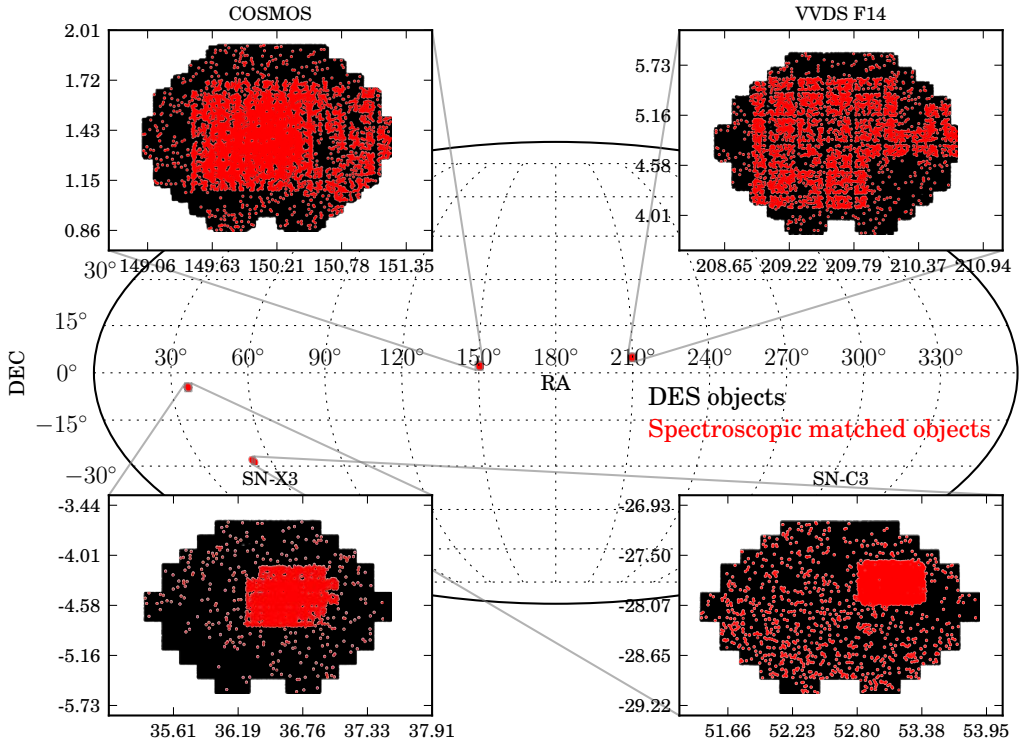

Figure 3.1: Positions in the sky of the four calibration fields. In the zoomed-in inset panels it is possible to observe the spectroscopic matched galaxies, in red, in front of all the DES galaxies detected in the fields, in black.

definition below in this section) at which the flux in a 2-arcsec diameter aperture is measured at 10σ. Note that for the SN-X3 and SN-C3 fields, we selected those SV observations that approximately met DES main survey sky background and seeing criteria in constructing the processed data used for this chapter.

The data were processed using the same routines used by DES Data Management (DESDM) in their processing pipeline (Mohr et al., 2012; Desai et al., 2012), in particular for image detrending, astrometric calibration (SCAMP, Bertin (2006)), image remapping and coaddition (SWarp, Bertin et al. (2002)), point spread function modeling (PSFEx, Bertin (2011)), and object detection and photometry (SExtractor, Bertin & Arnouts (1996)). The data were processed by running these codes in standalone mode at Fermilab, rather than by running them within the DESDM processing framework at NCSA. Running standalone was needed as the DESDM framework was not yet fully setup at the time (Spring 2013) to process and calibrate the data for these isolated fields all the way through to image coaddition.

Table 3.1: Imaging exposure times and depths for photo-z calibration fields.

Field	Filter	Tot. Exp. Time (sec)	10σ Depth
SN-X3 Main	u	900	22.6
SN-X3 Main	g	800	24.1
SN-X3 Main	r	1200	24.3
SN-X3 Main	i	1080	23.6
SN-X3 Main	z	990	22.7
SN-X3 Main	Y	500	21.0
SN-C3 Main	u	900	22.9
SN-C3 Main	g	800	24.3
SN-C3 Main	r	1200	24.3
SN-C3 Main	i	1080	23.6
SN-C3 Main	z	990	22.9
SN-C3 Main	Y	500	20.9
VVDS F14 Main	u	900	22.5
VVDS F14 Main	g	900	24.0
VVDS F14 Main	r	900	23.6
VVDS F14 Main	i	900	23.1
VVDS F14 Main	z	900	22.4
VVDS F14 Main	Y	750	21.6
SN-X3 Deep	u	900	22.6
SN-X3 Deep	g	2000	24.5
SN-X3 Deep	r	3600	24.9
SN-X3 Deep	i	5400	24.5
SN-X3 Deep	z	10890	24.0
SN-X3 Deep	Y	1000	21.0
SN-C3 Deep	u	900	22.8
SN-C3 Deep	g	1800	24.6
SN-C3 Deep	r	3600	24.9
SN-C3 Deep	i	5400	24.5
SN-C3 Deep	z	10890	24.3
SN-C3 Deep	Y	500	20.8
COSMOS Deep	u	33600	25.2
COSMOS Deep	g	4500	24.8
COSMOS Deep	r	4800	24.9
COSMOS Deep	i	12000	24.8
COSMOS Deep	z	7000	23.5
COSMOS Deep	Y	2400	21.6

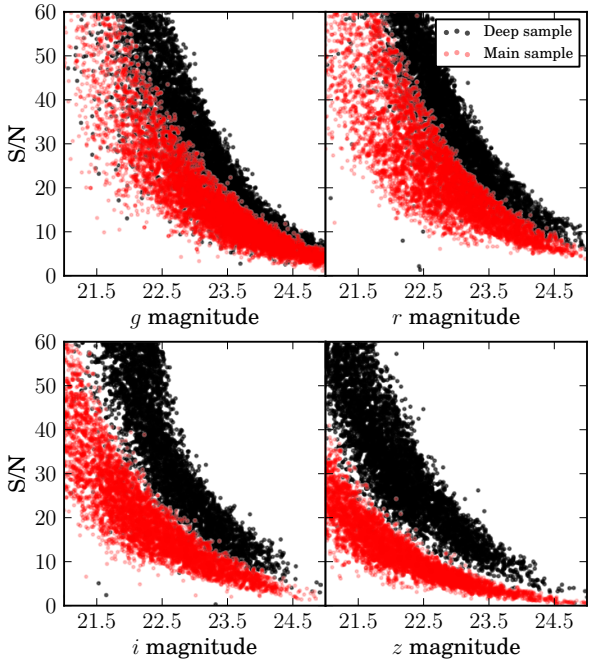

Figure 3.2: S/N vs. magnitude for g, r, i and z DES bands, and for Main (red dots) and Deep (black dots) samples.

Though we basically used the DESDM codes, there were some detailed differences in processing and photometric calibration that we highlight here. For image detrending we did not include corrections for CCD nonlinearity, pupil ghost, and illumination that are now used by DESDM, as these corrections were not available at the time. Image coaddition was done using a median coadd rather than by using a weighted mean as in DESDM. Photometric calibration in the $ugriz$ filters for the SN-X3, VVDS F14, and COSMOS fields was done by matching against overlapping bright stars from the SDSS Data Release 9 database (Ahn et al., 2012). This was done to calibrate each individual CCD on each separate DECam exposure, before image coaddition. In the Y band for all fields, and in all filters for the SN-C3 field (which did not overlap SDSS), we picked a fiducial exposure for each field, adopted the typical DECam CCD-by-CCD photometric zeropoints as determined from DES SV data, and then tied the photometry for subsequent exposures/CCDs to the fiducial exposure by matching overlapping bright objects. In addition, we also applied a further relative photometric calibration step, by selecting bright $r = 18$–22 galaxies in each field and offsetting the zeropoints in the other 5 filters so that the median galaxy colors relative to r (i.e., $g-r$, $r-i$, etc.) would match those for fiducial DECam data of the VVDS-Deep 02hr field (part of SN-X3). This additional step was intended to match up the median galaxy colors among the different fields by using photometry of galaxies directly, as the earlier calibration steps use photometry of stars, and there can be small (percent level) systematic differences between the stellar and galaxy photometric zeropoints, in particular due to seeing. We also applied corrections for Milky Way extinction based on the

Schlegel, Finkbeiner & Davis (1998) dust maps evaluated at the center of each field.

As for the use of imaging data by photo-z algorithms, either MAGAUTO or MAGDETMODEL magnitudes (or both) were employed by the different photo-z codes. MAGAUTO magnitudes come from the flux (counts) measured in an elliptical aperture defined as in Bertin & Arnouts (1996). It provides an estimation of the total magnitude of the object. MAGDETMODEL magnitudes are measured from the shape (a Sersic profile model (Sérsic, 1963)) fit to the object in the SExtractor detection image (either the r band or the i band for our data), and the flux is then measured separately in each band using that same model shape. Also available are MAGMODEL magnitudes, which fit the shape of the object independently in each of the bands. However, MAGDETMODEL magnitudes, which result from one unique best-fit shape for the object, are in general better suited for color measurement and hence more appropriate to use for photo-z estimation.

We want to emphasize here that because of the differences mentioned above between the reductions of SV data used in this chapter and the improved DESDM reductions of SV data (to be released and described elsewhere), the results of this chapter are meant to reflect the photo-z quality achievable from early DES data, rather than from final DES data or even from SV data. We expect that final DESDM reductions of the calibration field data will be better in terms of photometric quality and consequently of photo-z quality, so the results in this chapter will serve as a lower bound on the photo-z quality that may be achieved by final DES data. Nonetheless, as we will show later in this chapter, the photo-z quality achieved in these early DES data is good and already sufficient to meet the basic DES science requirements on photo-z scatter and outlier fractions.

3.3 DES-SV SPECTROSCOPIC SAMPLE

In general, to exploit a galaxy photometric survey to its maximum scientific potential, it is necessary to be able to calibrate or control the performance of the photo-z estimation by using data from a spectroscopic survey. To accomplish this, it is necessary to have, for a subset of galaxies, both the spectroscopic redshifts and the estimated photo-z's. With this information in hand, the characterization of the behavior of the photometric redshifts is possible, and it becomes a crucial step for cosmological probes such as galaxy clustering or weak lensing. In particular, among other quantities, it is very important to characterize the true redshift distribution of a set of galaxies after a selection in photo-z space.

A photometric survey like DES will therefore need to observe one or several regions of the sky that have been previously covered by a spectroscopic survey,

and then match the galaxies in the catalog of the spectroscopic survey to galaxies observed photometrically by DES. In this chapter, the matching between DES and spectroscopic galaxies is performed by using the positions of the galaxies in the sky plane, with a matching radius of 1 arcsec.

Four regions of the sky included in the DES-SV footprint have been used for photo-z calibration in this study (Fig. 3.1):

- **SN-X3 field**: This area, centered at RA $\sim$ 36, DEC $\sim$ -5, overlaps with the VIMOS (Le Fèvre et al. (2003)) VLT Deep Survey (VVDS) 02hr field. DES photometry has been matched in this field with spectroscopic redshift data from VVDS Deep (Le Fèvre et al. (2005, 2013)).
- **VVDS F14 field**: This area, centered at RA $\sim$ 209, DEC $\sim$ 5, overlaps with the VIMOS VLT Deep Survey (VVDS) 14hr field. DES photometry has been matched in this field with spectroscopic redshift data from VVDS Wide (Garilli et al. (2008)).
- **SN-C3 field**: This area, centered at RA $\sim$ 52, DEC $\sim$ -28, overlaps with the Chandra Deep Field South. DES photometry has been matched in this field with spectroscopic redshift data from both VVDS Deep and ACES (Cooper et al. (2012)).
- **COSMOS field**: This area, centered at RA $\sim$ 150, DEC $\sim$ -1.4, overlaps with the Cosmic Evolution Survey field. DES photometry has been matched in this field with spectroscopic redshift data from both VVDS Wide and zCOSMOS (Lilly et al. (2007, 2009)).

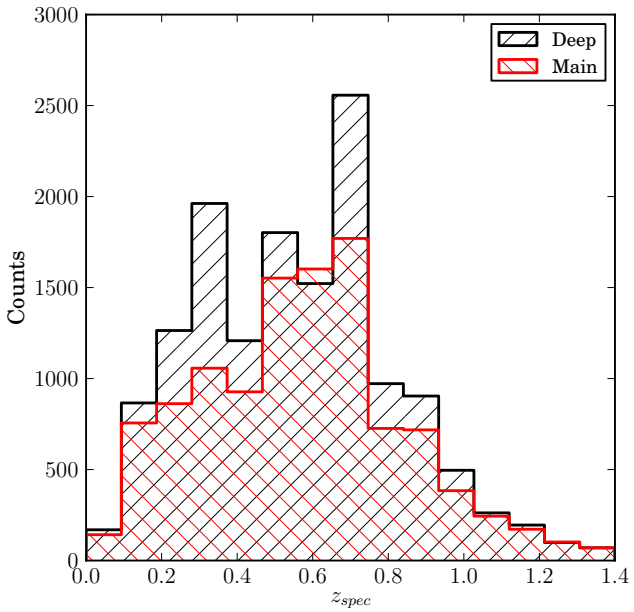

Figure 3.3: Redshift distributions for the spectroscopic matched galaxies in the Main and Deep calibration samples.

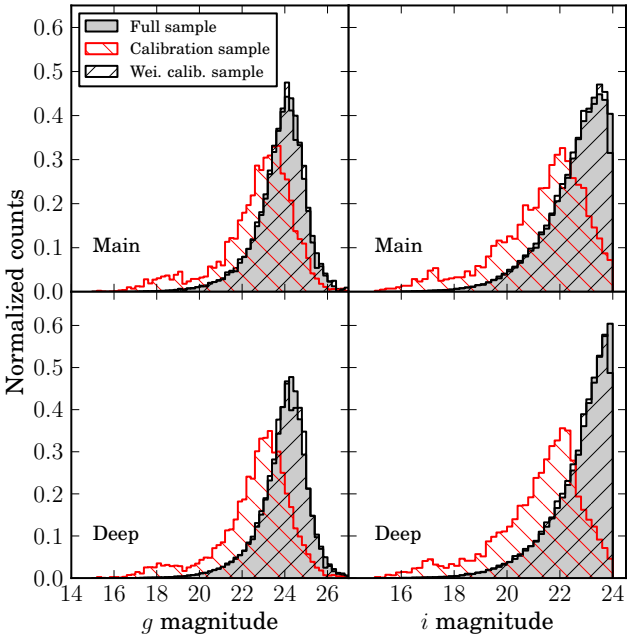

Figure 3.4: g and i magnitude distributions for the full, calibration and weighted calibration sample. The difference between the full and the calibration samples is apparent, the latter being significantly brighter. After applying the weighting procedure described in Lima et al. (2008), the weighted calibration distributions agree very well with the corresponding DES-SV distributions.

We have also used data from brighter, shallower spectroscopic surveys available in these four regions: OzDES (Kuehn et al. (2014), Yuan et al., in preparation); SDSS-I/II (Strauss et al., 2002; Eisenstein et al., 2001); SDSS-III BOSS Data Release 9 (Ahn et al., 2012) and 2dF Galaxy Redshift Survey (Colless et al., 2001). Galaxies matched to these surveys help to increase the statistics on the brighter tail of the distribution for galaxies used for this study.

Next we build a set of training and testing samples, using DES photometry from the Main and Deep samples and spectroscopy from different surveys. Note that only high-confidence spectroscopic redshifts, i.e., having redshift flags between 3 and 5, corresponding to secure and very secure ($>$ 95% accuracy) redshift determinations, have been selected to construct these samples. Spectroscopic failures can bias cosmological results, as studied in Cunha et al. (2012b), where they showed how for a final DES analysis such failures need to be reduced to the percent level. While a complete study concerning spectroscopic failures will have to take place for the analysis of the final DES data set, here we rely on the high-confidence redshift flags for the photo-z calibration of the early DES data. Below we describe how the data is distributed on each of the four calibration samples (training and testing for both Main and Deep catalogs):

- Main training sample: 5859 galaxies, photometry from Main catalogs, spectroscopic redshifts from the following data sets:

  - One randomly selected half of VVDS Deep, SDSS/BOSS in the SN-X3 field, ACES, 2dFGRS, OzDES in the SN-C3 field, and VVDS Wide, SDSS/BOSS in the VVDS F14 field.

- Main testing sample: 6381 galaxies, photometry from Main catalogs, spectroscopic redshifts from the following data sets:

  - The other half left out from the samples in the Main training set.

  - All VVDS Deep in the SN-C3 field.

- Deep training sample: 7249 galaxies, photometry from Deep catalogs, spectroscopic redshifts from the following data sets:

  - One randomly selected half of VVDS Deep, SDSS/BOSS in the SN-X3 field, ACES, 2dFGRS, OzDES in the SN-C3 field and zCOSMOS, SDSS/BOSS, 2dFGRS in the COSMOS field.

- Deep testing sample: 8358 galaxies, photometry from Deep catalogs, spectroscopic redshifts from the following data sets:

  - The other half left out from the samples in the Deep training sample.

  - All VVDS Deep in the SN-C3 field.

  - All VVDS Wide in the VVDS F14 field.

The spectroscopic redshift distributions of the Main and Deep calibration samples defined above, spanning all the redshift range of interest for DES ($0 < z < 1.4$), are shown in Fig. 3.3.

3.3.1 THE WEIGHTING PROCEDURE

In order to assess the photo-z performance of the DES-SV data we would ideally need a calibration sample being representative of the DES-SV full sample, i.e. having exactly the same photometric properties (magnitude and colour distributions). However, spectroscopic galaxy samples are shallower, and suffer from selection effects. A weighting procedure, which assigns a weight to each of the galaxies in the calibration sample so that the distributions of their photometric observables reproduce the distributions of the same observables in the full sample, can be used provided there is enough overlap between the photometric spaces of the calibration and full samples (Lima et al., 2008; Cunha et al., 2009).

Different algorithms can be used to compute the weights, but basically all compare local densities in the photometric spaces of the two samples (calibration and full) and assign a weight to each photometric region of the calibration sample equal to the ratio between the densities of galaxies in the full sample and the calibration sample in a given region. In this study we use a nearest neighbour algorithm to compute the weights that we use extensively throughout the chapter. A detailed description of the method can be found in Lima et al. (2008).

Table 3.2: Definition of the metrics used in the text to present the main results. These are computed in the whole redshift range as well as in bins of width 0.1 in photometric redshift. Detailed definitions can be found in the appendix.

Metric	Description	Req.
$\overline{\Delta z}$	mean of the Δz distribution	-
$\sigma_{\Delta z}$	standard deviation of the Δz distribution	-
Δz_{50}	median of the Δz distribution	-
σ_{68}	half width of the interval around Δz_{50} containing 68% of the galaxies	< 0.12
$\mathrm{out}_{2\sigma}$	fraction of galaxies with: $\lvert\Delta z - \overline{\Delta z}\rvert > 2\sigma_{\Delta z}$	< 0.1
$\mathrm{out}_{3\sigma}$	fraction of galaxies with: $\lvert\Delta z - \overline{\Delta z}\rvert > 3\sigma_{\Delta z}$	< 0.015
$\overline{\Delta z'}$	mean of the $\Delta z' = \Delta z/\epsilon_{\mathrm{phot}}$ distribution	-
$\sigma_{\Delta z'}$	standard deviation of the $\Delta z'$ distribution	-
$\mathrm{N}_{\mathrm{poisson}}$	difference between $N(z)^{\mathrm{phot}}$ and $N(z)^{\mathrm{spec}}$ normalized by Poisson fluctuations	-
KS	Kolmogorov - Smirnov statistic for $N(z)^{\mathrm{phot}}$, $N(z)^{\mathrm{spec}}$	-

We apply the weighting technique within a region in the multidimensional space defined by $18 < i_{AB} < 24$; $0 < g - r < 2$; $0 < r - i < 2$. In Fig. 3.4 one can check how the weighting procedure is efficiently applied for the sample used in this study. The figure shows, for two DES bands, and the Main and Deep samples, the magnitude distributions for the full sample, the calibration sample and the weighted calibration sample, whose distributions agree very well with those of the full sample.

3.4 PHOTOMETRIC REDSHIFTS IN THE DES-SV CALIBRATION SAMPLES

In this section we include all the photo-z analyses using the calibration data defined in Sec. 3.3. The analysis is carried out employing an extensive set of statistics. To construct most of the metrics used in this chapter, we first define the bias to be $\Delta z = z_{\mathrm{phot}} - z_{\mathrm{spec}}$ and the normalized bias by its reported error as $\Delta z' = (z_{\mathrm{phot}} - z_{\mathrm{spec}})/\epsilon_{\mathrm{phot}}$, where ϵ_{phot} is the error in the estimation of the photo-z's. We present the standard metrics used to compare the accuracy of the different codes in Table 3.2, together with the DES science requirements for photo-z's, set before the start of the survey. The DES science requirements are driven by the Dark Energy science that DES plans to carry out, in particular by weak lensing and large-scale structure tomographic measurements.

The photo-z metrics we consider are intended to measure the quality of the photometric redshifts in terms of their bias, scatter, and outlier fraction statistics, and also in terms of the fidelity of the photo-z errors and of the agreement between the photo-z and true redshift distributions. Detailed definitions of these metrics are given in Appendix A, while here we briefly summarize and motivate our choice

of metrics:

• *Bias:* To quantify the overall photo-z bias, we use the mean bias $\overline{\Delta z}$ and the median bias Δz_{50}.

• *Scatter:* To measure the photo-z scatter, we use both the standard deviation $\sigma_{\Delta z}$ of Δz and the 68-percentile width σ_{68} of Δz about the median (i.e., $\Delta z_{50} \pm \sigma_{68}$ covers 68% of the Δz distribution). For a Gaussian distribution of Δz, we would have $\sigma_{\Delta z} = \sigma_{68}$. However, in general Δz is not distributed as a Gaussian, so that σ_{68} measures the width of the core of the Δz distribution, whereas $\sigma_{\Delta z}$ is sensitive to the tails. The DES science requirements specify $\sigma_{68} < 0.12$ for a 90% of the selectable sample of galaxies.

• *Outlier Fractions:* To quantify the fraction of objects with large $|\Delta z|$, we measure the $2\sigma_{\Delta z}$ and $3\sigma_{\Delta z}$ outlier fractions $\text{out}_{2\sigma}$ and $\text{out}_{3\sigma}$, respectively, both defined relative to the mean photo-z bias $\overline{\Delta z}$. The DES science requirements limit these outlier fractions to be $\text{out}_{2\sigma} < 0.1$ and $\text{out}_{3\sigma} < 0.015$.

• *Fidelity of Photo-z Errors:* To assess the fidelity of the reported photo-z error ϵ_{phot}, we normalize Δz by ϵ_{phot} and calculate the resulting error-normalized mean bias $\overline{\Delta z'}$ and standard deviation $\sigma_{\Delta z'}$. Ideally, we would obtain $\overline{\Delta z'} = 0$ and $\sigma_{\Delta z'} = 1$, as for a Gaussian distribution of zero mean and unit variance. Deviations from these values would indicate inaccuracies in the photo-z errors.

• *Redshift Distributions:* Finally, to compare the photometric redshift distribution $N(z_{\text{phot}})$ to the true redshift distribution $N(z_{\text{spec}})$, we use two metrics. The first is N_{poisson}, which measures the rms difference between $N(z_{\text{phot}})$ and $N(z_{\text{spec}})$, normalized by Poisson fluctuations $\sqrt{N(z_{\text{spec}})}$. The second is the Kolmogorov-Smirnov (KS) metric that tests whether $N(z_{\text{phot}})$ and $N(z_{\text{spec}})$ are consistent with being drawn from the same parent distribution.

Some of the key DES science analyses, such as the galaxy angular correlation functions for large-scale structure studies, or cosmic shear tomography measurements for weak lensing and dark energy constraints, will use DES galaxies divided into separate photo-z bins. For these photo-z bins, the DES science requirements also specify stringent limits on the differences in bias, scatter, and outlier fractions between the DES photometric sample and the spectroscopic calibration sample. For example, there is a requirement that the mean bias $|\overline{\Delta z}| < 0.001(1+z)$ in bins of 0.1 in redshift. Accurate characterization of the full $P(z_{\text{spec}}|z_{\text{phot}})$ distribution, i.e., the distribution of true redshift in bins of photo-z, will also be needed for these DES science analyses. However, consideration of these more stringent DES photo-z science requirements is premature for the present chapter, because of the limited number and sky distribution of the SV spectroscopic calibration fields. These fields are subject to sample variance effects, i.e., fluctuations in galaxy densities and redshift distributions due to clustering and large-scale structure, and in

fact, as detailed in Cunha et al. (2012a), meeting the requirements will necessitate a much larger number of widely-distributed spectroscopic calibration fields (e.g., ∼100) than are currently available from completed redshift surveys. We will thus postpone discussion of these issues and tests of these more stringent requirements for future DES photo-z papers. Nonetheless, we do present in a later section some example $P(z_{\rm spec}|z_{\rm phot})$ distribution for several selected photo-z codes.

All the results shown in this chapter have been weighted according to the technique presented in Section 3.3.1, and they include a cut on the 10% of the galaxies having larger estimated photo-z error, as given from each particular code (this 10% cut on photo-z error is allowed by the DES science requirements). This cut introduces small differences among the testing samples used by each photo-z code in the comparison. With this we test the ability of each particular code to select the most problematic objects in the sample. In addition, clustering measurements can be affected by severe photo-z quality cuts as presented in Martí et al. (2013), where they also show a procedure to correct for these effects.

All the errors for the statistics presented in the chapter come from bootstrap resampling using 100 samples, unless otherwise stated.

3.4.1 METHODS

Before going in detail into the photo-z analysis we present here a brief description of the different methods we have used to estimate photo-z's, which include most of the relevant photo-z codes available. We have emphasized in the details at the time of running these codes; for an exhaustive description of them see Table 3.3. For template-based methods, a standardized set of filter throughput curves has been used. Most of the codes have been run in standalone mode, while a fair fraction of them has been run within the DES Science Portal, with compatible results. Due to the large number of codes used, the chapter, other than showing the DES-SV photo-z capabilities, also serves as a helpful reference to compare different photo-z codes using real data from a deep galaxy survey.

DESDM

The DESDM (default) photo-z's were computed using the same artificial neural network method that was applied to the Sloan Digital Sky Survey Data Release 6 (DR6) sample, as described in detail by Oyaizu et al. (2008b). In brief, we used a neural network configuration with 10 input nodes, consisting of the 5 $grizY$ `MAGAUTO` magnitudes and the 5 $grizY$ `MAGDETMODEL` magnitudes, followed by 3 hidden layers with 15 nodes per layer. The formal minimization to determine the neural network weights was done on the training set data, while choosing

Table 3.3: List of methods used to estimate photo-z's. Code type and main references are given.

Code	Type	Reference
DESDM, Artificial Neural Network	Training	Oyaizu et al. (2008b)
ANNz, Artificial Neural Network	Training	Collister & Lahav (2004)
TPZ, Prediction Trees and Random Forest	Training	Carrasco Kind & Brunner (2013, 2014)
RVMz, Relevance Vector Machine	Training	Tipping (2001)
NIP-kNNz, Normalized Inner Product Nearest Neighbor	Training	de Vicente et al., in preparation
ANNz2, Machine Learning Methods	Training	Sadeh et al., in preparation
ArborZ, Boosted Decision Trees	Training	Gerdes et al. (2010)
SkyNet, Classification Artificial Neural Network	Training	Bonnett (2013); Graff et al. (2013)
BPZ, Bayesian Photometric Redshifts	Template	Benitez (2000); Coe et al. (2006)
EAZY, Easy and Accurate Redshifts from Yale	Template	Brammer, van Dokkum & Coppi (2008)
LePhare	Template	Arnouts et al. (2002); Ilbert et al. (2006)
ZEBRA, Zurich Extragalactic Bayesian Redshift Analyzer	Template	Feldmann et al. (2006)
Photo-Z	Template	Bender et al. (2001)

the set of network weights that gave the lowest photo-z scatter on the testing set, after 300 iterations of the weight minimization. Moreover, to reduce sensitivity to initial conditions in the minimization procedure, we repeated the procedure 10 times, starting each time at a different initial position in the space of weights. The final photo-z for a galaxy was taken to be the average of the photo-z's computed from the optimal weights for each of the 10 network minimizations.

We also computed accompanying photo-z errors using the empirical *nearest neighbor error* (NNE) technique, described in detail by Oyaizu et al. (2008a). The NNE method estimates the photo-z error for each galaxy empirically, based on the photo-z's and true redshifts of the galaxy's 100 nearest neighbors in the spectroscopic testing set, where neighbor distance is defined using a simple flat metric in the space consisting of the 10 input magnitudes noted above. Specifically, the NNE photo-z error σ is defined so that it corresponds to the width of 68% of the $|z_{phot} - z_{spec}|$ distribution of the nearest neighbors.

ANNz

ANNz (Collister & Lahav (2004)) is a training-based method that uses a neural network scheme to find a functional relationship between a given set of input parameters (e.g. magnitudes, colors, etc) and outputs a desired quantity (e.g. redshift). The results shown in this chapter have been obtained by using a neural network architecture of 2 layers with 10 nodes each, and using as inputs the set of 5 `MAGAUTO` and 5 `MAGDETMODEL` magnitudes. Attempts to use a larger number of nodes as well as colors as inputs resulted in larger photo-z errors. The uncertainties in the photo-z estimation are computed using standard propagation of the errors in the input magnitudes to the error in photo-z, by using the functional relationship between these input parameters and the output photo-z.

TPZ (TREES FOR PHOTO-Z)

TPZ[3] (Carrasco Kind & Brunner, 2013, 2014) is a machine learning, parallel algorithm that uses prediction trees and random forest techniques to produce both robust photometric redshift pdfs and ancillary information for a galaxy sample. A prediction tree is built by asking a sequence of questions that recursively split the input data taken from the spectroscopic sample, frequently into two branches, until a terminal leaf is created that meets a stopping criterion (e.g., a minimum leaf size or a variance threshold). The dimension in which the data are divided is chosen to be the one with highest information gain among the random subsample of dimensions obtained at every point. This process produces less correlated trees and allows the exploration of several configurations within the data. The small region bounding the data in the terminal leaf node represents a specific subsample of the entire data with similar properties. Within this leaf, a model is applied that provides a fairly comprehensible prediction, especially in situations where many variables may exist that interact in a nonlinear manner as is often the case with photo-z estimation.

By perturbing the data using their magnitude errors and by taking bootstrapping samples, many (600 in this application) uncorrelated trees can be created whose results are aggregated to construct each individual pdf. For the application to DES-SV data, we have used both `MAGAUTO` and `MAGDETMODEL` magnitudes in the five DES bands, together with all the corresponding colors as well as their associated errors.

RVMZ

RVMz is an empirical photo-z code based on the relevance vector machine algorithm from Tipping (2001), a Bayesian sparse kernel method for regression. The relevance vector machine (RVM) has characteristics similar to the support vector machine, but includes a Bayesian treatment for the determination of the model weights. This has the advantage that the parameters governing model complexity and noise variance are found in the training run itself, and therefore the RVM does not require cross validation to optimize these parameters. We use the RVM implementation in the R-package `kernlab` from Karatzoglou et al. (2004). To obtain photo-z estimates, we used `MAGDETMODEL` magnitudes ($grizY$) and colors ($g-r$, $r-i$, $i-z$, $z-Y$) as input. We reconstruct the pdf by combining the uncertainties in the datasets and the model. In the training set we use the k-fold cross validation technique, which consists of partitioning the data in k groups, then $k-1$ of these groups are used to train the model which is then evaluated on

[3] http://lcdm.astro.illinois.edu/research/TPZ.html

the hold-out group. This process is then repeated for all possible choices of the hold-out group and the resulting mean squared error for the redshift prediction is evaluated. At this stage we obtain the model error as the RMS of the predicted pdf. Details of the method will be described in Rau et. al. (in prep.).

NIP-kNNz

NIP-kNNz (Juan De Vicente et al. (2014), in preparation) is a novel technique that computes the photo-z from a Nearest Neighbour approach based on the Normalized Inner Product (NIP). While Euclidean magnitude-distance ensures that close galaxies in magnitude space are assigned the same redshift, it does not considers as neighbors galaxies with the same color but separated in overall magnitude. NIP metrics corrects this by considering two galaxies as neighbors, and hence with close redshift, when they have similar colors, rather than magnitudes. The metric is based on the inner product definition:

$$\mathrm{NIP} = \cos\alpha = \frac{\mathbf{M}_t \cdot \mathbf{M}_p}{M_t M_p} \ , \tag{3.1}$$

where $\mathbf{M}_t$ and $\mathbf{M}_p$ are the multi-magnitude vectors of training and photometric galaxies respectively. For this particular application, the five `MAGDETMODEL` magnitudes were used and turned into fluxes. The normalized inner product is related to the angle that the two multi-magnitude vectors form. Maximizing NIP is equivalent to minimize the angle between the two vectors. Regarding the photo-z error, an empirical formula has been derived to account for three different contributions. The first term is the floor error related to the finite spectroscopic redshift precision, which is all that remains In the best case scenario, when the magnitude vectors of the photometric and the spectroscopic galaxies point in the same direcction. The second contribution comes from the characterization of the photo-z errors in the spectroscopic sample. NIP-kNN is run over all galaxies in the spectroscopic sample to obtain their photo-zs. One half of the difference between the spectroscopic z and the photo-z of the spectroscopic galaxy is taken as its photo-z error. When NIP-kNN is applied to a galaxy in the photometric sample, it inherits not only the z of the closest spectroscopic galaxy but also its photo-z error. The third term is the metric distance $\sin(\alpha)$ that accounts for neighborhood, multiplied by a constant determined empirically. Assuming the spectroscopic sample spans the range of redshifts of the photometric sample, NIP-kNNz achieves, by construction, an accurate reconstruction of the redshift distribution $N(z)$

ANNz2

ANNz2 is a new major version of the public photo-z estimation software, ANNz (Collister & Lahav, 2004), which will be made public in 2015. The new code incorporates several machine-learning methods, such as artificial neural networks (ANNs), boosted decision trees (BDTs, Freund & Schapire (1997), described below in 3.4.1) and k-nearest neighbors (KNNs). The different algorithms are used in concert in order to optimize the photo-z reconstruction performance, and to estimate the uncertainties of the photometric solutions. This is done by generating a wide selection of machine-learning solutions with e.g., different ANN architectures, initialized by different random seeds. The optimization is performed by ranking the different solution according to their performance, which is determined by the respective photo-z scatter of each solution.

The single solution with the best performance is chosen as the nominal photo-z estimator of ANNz2. In addition, the entire collection of solutions is used in order to derive a photo-z probability density function (pdf), constructed in two steps. First, each solution is folded with an error distribution, which is derived using the KNN error estimation method of Oyaizu et al. (2008b). The ensemble of solutions is then combined. This is done by weighing the different estimators, in such a way as to produce a pdf which describes the underlying photometric errors. The inputs used in this study were the five `MAGAUTO` and the five `MAGDETMODEL` magnitudes.

ARBORZ

The ArborZ algorithm (Gerdes et al., 2010; Sypniewski, 2014) is a training-set-based, publicly-available (Sypniewski & Gerdes, 2013) photo-z estimator that makes use of boosted decision trees (BDTs). BDTs were developed to classify objects characterized by a vector of observables $\mathbf{x}$ into two categories. Decision trees are trained iteratively, with initially misclassified objects given higher weight, or "boosted", in the next training cycle. An individual decision tree is a relatively weak classifier. But the "forest" of trees generated during the training process, when their outputs are combined in a way that assigns higher weight to trees with lower misclassification rates, collectively constitutes a strong classifier. To adapt a binary classifier to the problem of determining a continuous quantity like redshift, we divide the redshift range of interest into N discrete bins with a width roughly 25-50% of the expected photo-z resolution, and train a separate BDT classifier for each redshift bin, using a forest size of 50 trees. Each classifier is trained to identify galaxies with a redshift falling in its particular bin as "signal," and to reject galaxies falling more than $3\sigma_{z,phot}$ away from its bin as "background." The

$3\sigma_{z,phot}$ exclusion region between signal and background objects is introduced in order to avoid the overtraining that could result from treating a galaxy with, e.g., a redshift of 0.999 as signal and one with 1.001 as background. Each BDT classifier in this ensemble, when presented with a new galaxy, produces a probability that the given galaxy falls within its redshift bin. This collection of probabilities constitutes the ArborZ pdf. ArborZ also provides a single best-estimate photo-z (from the median of the pdf) and its error, σ_{68} (from its width); however, the full pdf provides the best characterization of a galaxy's photo-z.

In the DES-SV sample, we train ArborZ using the `MAGAUTO` and `MAGDETMODEL` magnitudes in $grizY$, with 50 fixed-width redshift bins out to $z = 1.5$. The results are fairly robust with respect to reasonable variations in the forest size, number of bins, the choice of fixed- or variable-width bins, and the use of either or both sets of magnitudes.

SKYNET, CLASSIFICATION NEURAL NETWORK

This method, first used in Bonnett (2013), consists of using a neural network to classify galaxies in classes, in this case redshift bins. A neural network with a $softmax$ transformed output (Pyle, 1999) is able to estimate the probability that an object belongs to a certain class. Given the fact that a galaxy cannot live in more than one redshift bin at the same time, a neural network with a *softmax* transformation is ideally suited to return a pdf for each galaxy. Before training the neural network, we bin our data in n redshift bins, the classes. The neural network is fed the `MAGAUTO` magnitudes, the `MAGDETMODEL` magnitudes and the correct classes. The neural network outputs n values between $[0, 1]$ for each galaxy. These n values sum up to 1 and hence can be interpreted as the probability that a galaxy resides in a redshift bin. In this particular run we used $n = 40$, resulting in a $\delta z = 0.035$ granularity in the probability density function. The publicly available neural network software SkyNet (Graff et al., 2013) was used for this work. The neural net is trained using 3 hidden layers with respectively 20, 30 and 40 nodes per layer.

BPZ

The BPZ (Bayesian Photometric Redshifts) photo-z code from Benitez (2000) and Coe et al. (2006) is a template-based method that returns the whole probability density distribution $p(z|m_i)$ that the galaxy is at redshift z when its magnitudes in each band are m_i, and also a single photo-z value computed as the maximum of $p(z|m_i)$. Following Bayes Theorem, $p(z|m_i)$ is the product of a likelihood and a prior probability function that represents our previous knowledge of the redshift

and spectral type t distributions of the sample in the analysis. In the likelihood, we use the five MAGDETMODEL magnitudes.

- *Templates:* We use the eight spectral templates that BPZ carries by default based on Coleman, Wu & Weedman (1980); Kinney et al. (1996), and add two more interpolated templates between each pair of them by setting the input parameter INTERP=2 (option by default).

- *Prior:* We explicitly calibrate the prior in each test by fitting the empirical function $\Pi(z, t|m_0)$ proposed in Benitez (2000) to the corresponding DES-SV training sample. With this, we are able to remove most of the catastrophic outliers which for template-based methods tend to constitute a sizable fraction of all galaxies.

- *Training:* No other training or calibration has been attempted.

EAZY

The EAZY photo-z code (Brammer, van Dokkum & Coppi, 2008) is a template-based maximum likelihood method that has been specifically optimized for use when representative spectroscopic redshifts are not available for training set based estimators. In this implementation we use the five MAGDETMODEL magnitudes.

- *Templates:* The code uses a novel non-negative matrix factorization algorithm to construct a minimal set of templates which are linear combinations of templates derived from semi-analytic models, based on the Bruzual & Charlot (2003) models but with the star formation histories computed from the semi-analytic models of De Lucia & Blaizot (2007). These model templates are likely to be more representative of galaxies, particularly at high redshifts, compared to the commonly used Coleman, Wu & Weedman (1980) or Bruzual A. & Charlot (1993) templates. The code also makes use of a template error function to account for uncertainties in templates over specific wavelength ranges.

- *Priors:* No priors are used.

- *Training:* No calibration of the templates is performed using spectroscopic training data. Equally, photometric offsets cannot be derived using spectroscopic training sets. In instances where significant offsets need to be applied to the photometric catalogues, the code is therefore unlikely to perform well relative to template-based codes where these offsets can be directly estimated from the training data.

LEPHARE

LePhare (Arnouts et al. (2002); Ilbert et al. (2006)) is a public template fitting code that uses a χ^2 minimization of differences between observed and theoretical magnitudes to find the best template (galaxy type) and redshift. The code also provides for each object upper and lower 1-sigma limits for this estimate, a maximum likelihood estimate for the redshift, K-corrections and absolute magnitudes for each band and a probability distribution function. An adaptive method can be used to improve theoretical magnitudes, as well as an $N(z)$ prior to minimize catastrophic errors. The effect of emission lines on the theoretical magnitudes can be estimated and taken into account. Several sets of SEDs and extinction laws are available in the code to be used. For this study we use the five `MAGDETMODEL` magnitudes.

- *Templates:* A set of 66 SEDs were used in the analysis of the CFHTLS data by Ilbert et al. (2006) obtained from interpolation of the largely used Coleman, Wu & Weedman (1980) templates for different Hubble types and Kinney et al. (1996) for starburst galaxies. Since template fitting codes are time consuming, we searched for a reduced group of templates from this large set, yielding essentially the same overall statistics as far as dispersion and outlier fraction are concerned. We performed several tests using the VVDS-02hr sample with available spectroscopic redshifts and found a reduced set of 21 templates encompassing SEDs for 12 Ellipticals, 6 Spirals, 1 Im and 2 starburst with satisfactory results. Several tests removing the u-band from VVDS and CFHTLS data indicated that the discrepancies of photo-z from the true (spectroscopic) value increase due to galaxy type–extinction degeneracy. From these tests we concluded that in order to minimize this problem we should keep the extinction values E(B-V) less or equal to 0.25 (for types Scd and later) and use only three very late type SEDs (1 Im and 2 starbursts).

- *Prior:* Although the $N(z)$ prior available in LePhare refers to the (B-I) color from the VVDS survey, we verified that applying the procedure to the $g-i$ DES color we achieved a significant minimization of the outlier fraction, and therefore we used this prior.

- *Training:* The adaptive training method available in LePhare was used to obtain a re-calibration (zero-points offsets) in each band. This procedure was first applied to the training sets and the shifts obtained were used when the code was run on the testing samples.

ZEBRA

The Zurich Extra-galactic Bayesian Redshift Analyser (Feldmann et al. (2006)) is a flexible and powerful photometric redshift code, based around template fitting. The code produces a posterior distribution for each galaxy in redshift and template space, $P(z, T)$, as well as marginalized distributions for $P(z)$ and $p(T)$. For redshift computation, the filters were smoothed over a scale of 100, and the templates smoothed over 30. The tests were conducted with a redshift resolution in linear steps of $\Delta z = 0.01$. The five `MAGDETMODEL` $grizY$ magnitudes were used in this analysis.

- *Templates:* 81 templates were used in ZEBRA's Bayesian mode. These templates were selected from a super-set of SEDs, and consist of the most frequent best-fit templates at $z = z_{\rm spec}$ for galaxies in the training sample. The super-set of template SEDs were produced by ZEBRA's template correction module from log interpolations between the Coleman, Wu & Weedman (1980) and Kinney et al. (1996) templates. The module allows the user to define redshift intervals within which the templates are modified to better fit the input photometry. For this stage a photometric sample from the COSMOS field was used. A fraction of the testing set galaxies have counterparts within this photometric sample, but the photometry is independent from the DES-SV data and the specroscopic redshifts were not used.

- *Prior:* In Bayesian mode, ZEBRA constructs a self-consistent iterative prior from the galaxy likelihood functions, $L(z, T)$. The approach is similar to that taken by Brodwin et al. (2006), but operates in 2-dimensional redshift and template space. The prior constructed from the training sample was used for both samples.

- *Training:* Bordoloi, Lilly & Amara (2010) describe a method of using known (spectroscopic) redshifts to correct the individual marginalized redshift probability distributions, $P(z)$. The method demands that the spectroscopic redshifts sample their respective $P(z)$ fairly, i.e. the distribution of cumulative probabilities between zero and $z_{\rm spec}$ should be flat. We apply a simple first pass of their approach in bins of redshift, with width $\Delta z_{\rm phot} = 0.1$. Galaxies were assigned to these bins based on their maximum posterior redshift. After correction of the individual $P(z)$, a new $z_{\rm phot}$ was computed as the maximum of the corrected $P(z)$.

PHOTOZ

PhotoZ (Bender et al., 2001) is a Bayesian template fitting photometric redshift code. The redshift probability of an object is obtained by multiplying the probability of a χ^2 fit of template SEDs by prior probabilities for redshift and luminosity. The total probability of a model then reads:

$$P(\vec{\mu}|m) \propto \mathscr{L}(m|\vec{\mu}) \cdot P(\vec{\mu}), \tag{3.2}$$

where m denotes the photometric data (in magnitudes or fluxes), and $\vec{\mu}$ are the model parameters, i.e., redshift z and luminosity M. In this analysis we used the five MAGDETMODEL $griz Y$ magnitudes.

- *Templates:* The template set we use contains templates ranging from star-forming (blue) to passively evolving (red) galaxies. It includes model SEDs from Bender et al. (2001), which were created from spectroscopically observed objects from the Hubble Deep Field North. Another three templates (an S0, Sac, and an Sbc galaxy) are from Mannucci et al. (2001), and two empirical SEDs (of an Scd and an Sbc galaxy) are from Coleman, Wu & Weedman (1980). Our model set additionally includes 13 SEDs from Ilbert et al. (2006) which are based on spectra from Coleman, Wu & Weedman (1980) and were optimized to match local star-forming galaxies. This is a combination of template sets already used in the past for photometric redshift estimation (e.g., Bender et al. 2001 and Brimioulle et al. 2013). Furthermore, we incorporate a set of red SEDs in our model set which were created in order to match the colors of luminous red galaxies (LRGs) from SDSS-II (Greisel et al., 2013).

- *Priors:* The redshift and luminosity priors have the form $P(x) \propto \exp(-((x-\hat{x})/\sigma)^p)$, where $\hat{x}$, σ, and p are defined individually for each model SED. Setting $\hat{x}$, σ, and p accordingly, we can, for instance, decrease the probability of observing red models at higher redshifts ($z \gtrsim 0.9$), or of galaxies that are too bright in absolute magnitude to exist. In addition to that, we adapt the z and M (absolute magnitude) priors for every model SED in such a way that photometric redshift outliers with $|z_{phot} - z_{spec}|/(1+z) > 0.15$ (Ilbert et al., 2006) in the Main and Deep DES-SV training sets are less likely. Therefore, we identify their location in the z vs. M space and modify the priors in such a way that they assign smaller probabilities to those regions. This is done solely if the outliers of a template are isolated from good photometric redshift estimates of the same template in the z vs. M space.

- *Training:* We iteratively adapt the zero-points for the training catalogs using the median magnitude offsets between the data and the model predictions while optimizing the photometric redshift performance.

3.4.2 RESULTS OF THE PHOTO-z ANALYSES

After the description of the codes, we turn to the study of their performance in a number of tests using different configurations of the data samples. While most of these tests focus on estimating how the photo-z determination will perform for the standard DES data, others look for improvements by using deeper photometry or additional bands. We also check the differences in the results under variations in the calibration data and the weights used. Note that the results presented in this subsection are those considering all the galaxies (with quality cuts), which are represented by one single statistic, later in the chapter we analyze some of these results in more detail.

TEST 1: *Main-Main*

This test is the most representative of the results shown in this chapter, the default case. We use here the Main training sample to train and calibrate the photo-z algorithms and the Main testing sample to validate them, therefore, the test represents the real situation for most of the data collected in the DES survey.

In order to display the performance of all codes, and only for this test in particular, in Fig. 3.5 we show the z_{phot} vs. z_{spec} scatter plot for all the codes listed in Table 3.3. Furthermore, we compute all the metrics presented in Table 3.2 and described in Appendix A. The results, using all the objects in the testing sample except for the 10% quality cut mentioned above, are shown in Table 6 and Figs 3.6-3.9. The legend is only shown in Fig. 3.6, but applies to subsequent figures corresponding to this test.

Figure 3.6 shows σ_{68}, related to the precision of the photometric redshifts (and defined in Appendix A), versus the mean bias of the photo-z's. The black dashed line sets the DES science requirement on σ_{68}, and one can check how most of the codes presented in this work are below this line, thus fullfilling this important requirement on precision. Also, among the codes satisfying the σ_{68} requirement, there is a subgroup having very low bias as well. In Fig. 3.6 we show a zoomed-in of this region of interest, where we can see how training-based codes, either producing a single photo-z estimate or a probability density function, $P(z)$, are the ones showing best performance (all the codes in the zoomed-in region belong to the training-based category).

Figure 3.7 shows the 3σ vs. 2σ outlier fractions for Test 1. The requirement

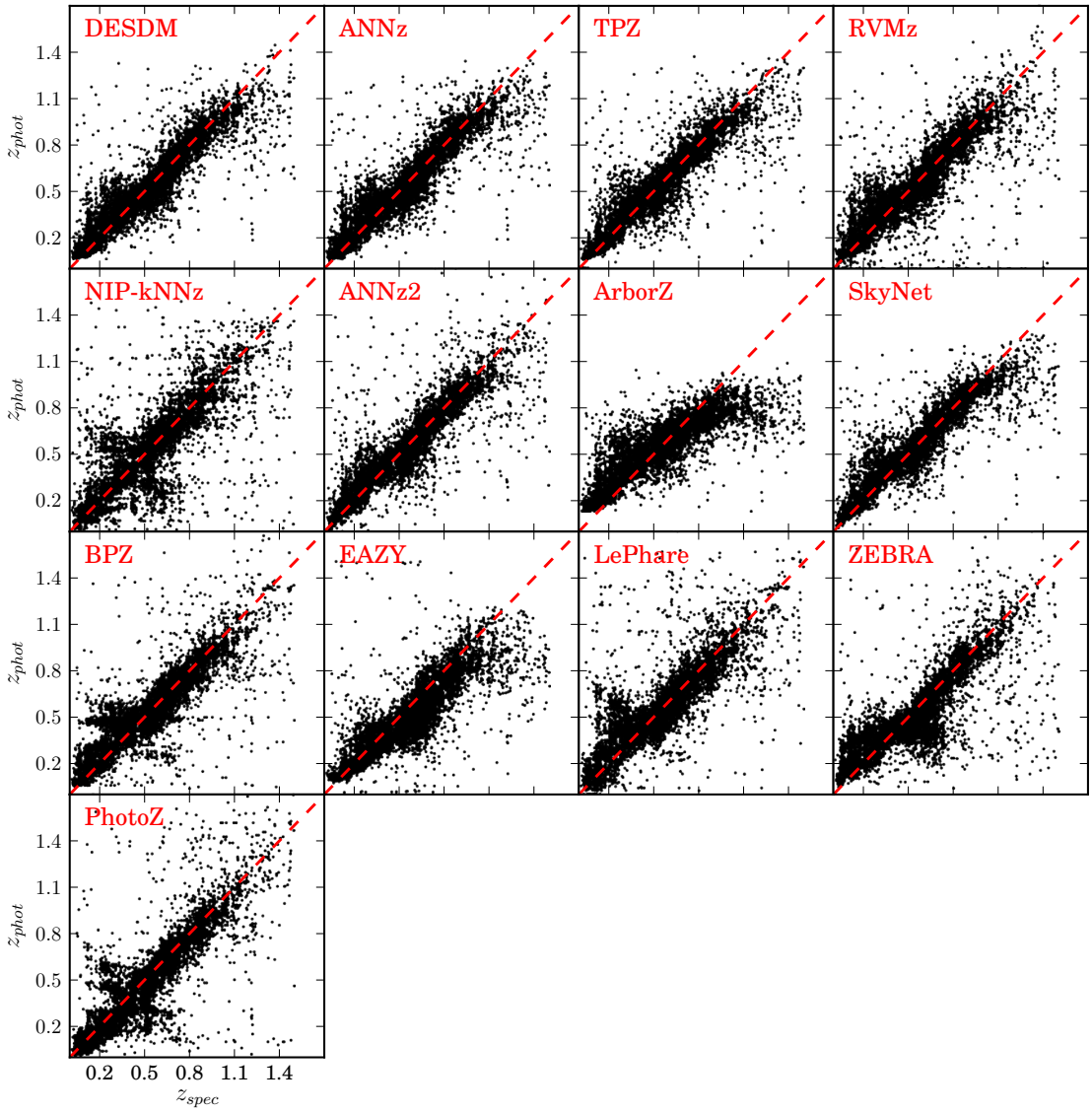

Figure 3.5: z_{phot} vs. z_{spec} scatter plot for all the codes analyzed in Test 1 and listed in Table 3.3.

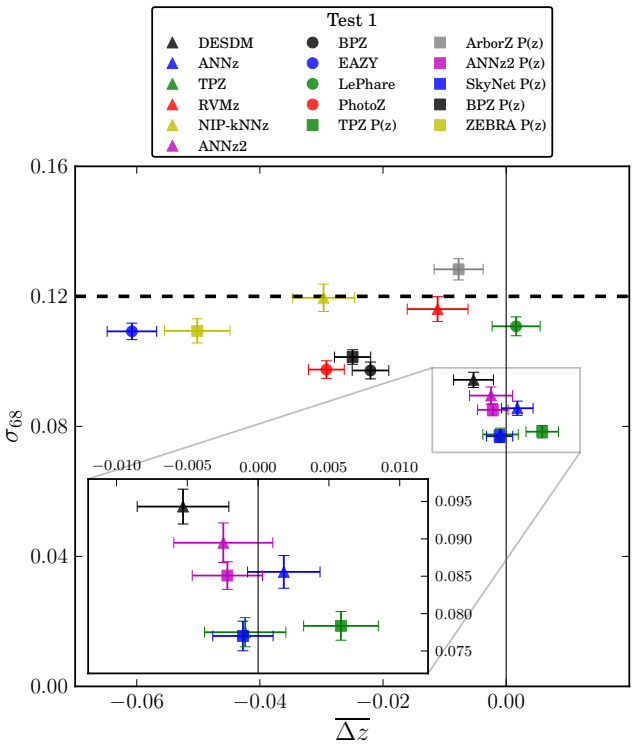

Figure 3.6: σ_{68} vs. bias for all the codes analyzed in Test 1. Black dashed lines represent the DES science requirements in this and subsequent figures. Training-based codes have triangles as markers, template-based have circles, and codes producing a probability density function (pdf) for the redshift are marked with a square. This will also be the convention for the next figures. Training-based codes, either producing a single photo-z estimate or a pdf, are the only ones present in the region of best performance (zoom-in).

on the 2σ outlier fraction (0.1) is beyond the range of the plot, meaning that all codes fullfill this. However, the 3σ outlier fraction requirement, shown as the black dashed line, is only met by a few codes. Among these codes, there are cases from the two types of photo-z codes, training and template-based. Also, there is more homogeneity in this plot: many codes agree with others within error bars. Both of these requirements are set based on the spread of the Δz distribution and no with respect to a fixed distance from the mean, therefore these values quantify how sharp the Δz distribution is with respect to its center.

Figure 3.8 shows RMS vs. bias for the Δz distribution normalized by its error ($\Delta z'$) of each of the codes analyzed in Test 1 ($\sigma_{\Delta z'}$ vs. $\overline{\Delta z'}$ in our notation). A large fraction of the codes yield very high values of $\sigma_{\Delta z'}$ (expected to be close to 1), meaning that all these codes underestimate their photo-z errors, however, there is a group of codes with normalized Δz distributions approaching a Gaussian with mean equals to zero and variance equals to one. Also, we do not see a particular type of photo-z code being problematic here, both training and template-based codes populate good and bad regions of the plot.

One crucial aspect of photo-z studies, which will be discussed in more detail in Section 3.4.3, is the estimation and calibration of the true galaxy redshift distributions. In this chapter we use two metrics to compare the reconstruction of the true redshift distribution by the different photo-z algorithms: the $N_{poisson}$ and KS statistics, defined in Appendix A. In both cases, the smaller the value, the

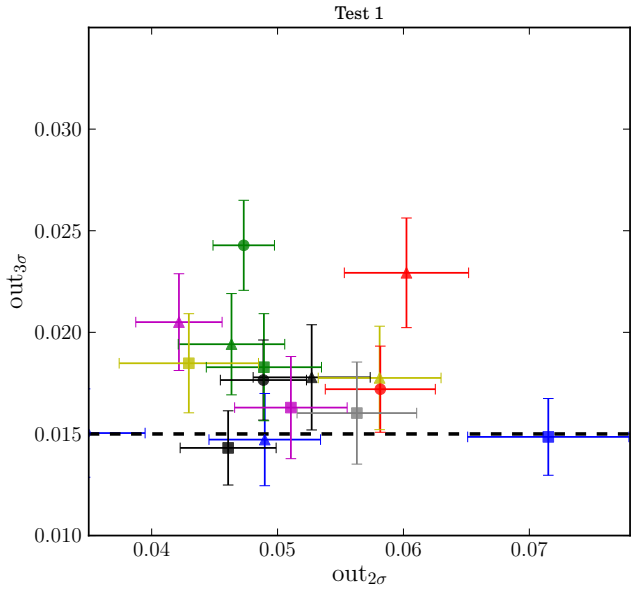

Figure 3.7: 3σ vs. 2σ outlier fraction for all the codes analyzed in Test 1. The results are more homogeneous than those for σ_{68} vs. bias (Fig. 3.6).

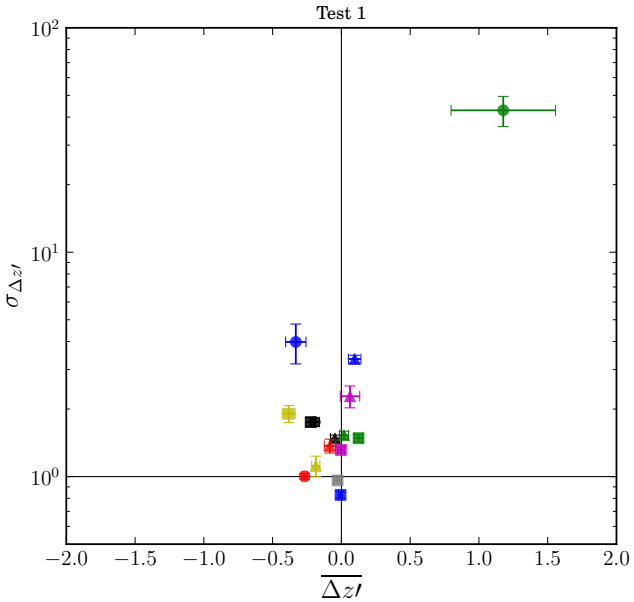

Figure 3.8: RMS vs. bias of the normalized Δz distribution $\Delta z' = \Delta z/\epsilon_{\mathrm{phot}}$ for all the codes analyzed in Test 1. Most codes underestimate ϵ_{phot}, leading to large values of $\sigma_{\Delta z\prime}$.

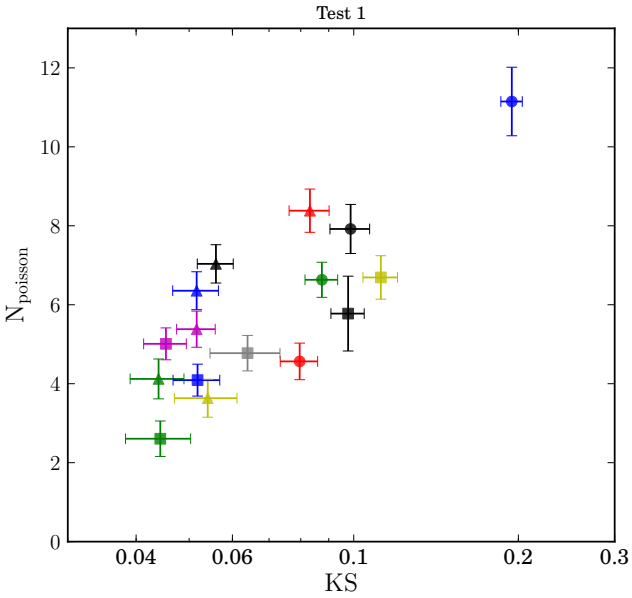

Figure 3.9: $N_{poisson}$ vs. KS statistics for all the codes analyzed in Test 1. Both metrics show how the true galaxy redshift distribution is reconstructed through photo-z's, for each code. The smaller the value of the metric, the better the reconstruction. A strong correlation between the two metrics is observed, as expected.

closer are the true redshift distribution and its reconstruction through photo-z's. Figure 3.9 shows these values for all the codes analyzed in Test 1. As expected, the two metrics are strongly correlated. It can also be seen how having a redshift pdf for each galaxy, instead of a single-estimate photo-z, helps a given code to have a better redshift reconstruction. This can be inferred looking at the cases where both the pdf and the single-estimate are displayed (TPZ, ANNz2, BPZ): in all these cases the pdf version of the code obtains better results in terms of these two metrics. As for the results, TPZ and the nearest-neighbor code, NIP-kNNz, show the best performance in this regard.

To summarize the results from Test 1, we note that most of the codes presented in this work fulfill the requirements for σ_{68} and 2σ outlier fraction, while only a few fulfill the 3σ outlier fraction requirement. Also, training-based codes seem to yield better photo-z precision on average and better $N(z)$ reconstruction, but, when evaluating other quantities like outlier fraction or the estimation of photo-z errors, there is no a clear indication as of which class of photo-z approach show more accurate metrics. As pointed out in Carrasco Kind & Brunner (2014) these results might vary for different regions on the multidimensional photometric space or within the redshift range. Usually, training-based algorithms perform better on areas well populated with training galaxies and poorly on those less dense regions (as in high redshift bins), fact that we can observe from Figure 3.5 where training-based methods tend to have tighter distributions at the center while some template-based methods can compute photo-z's for galaxies at higher redshift more efficiently.

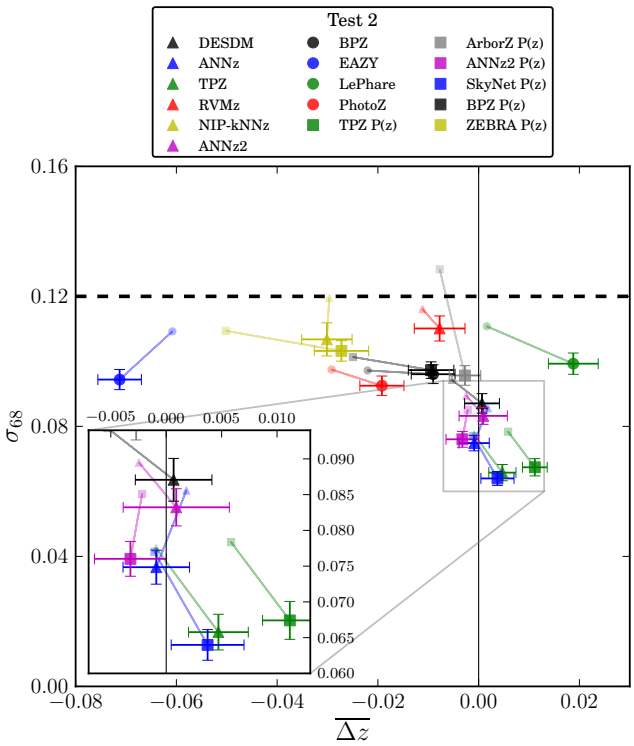

Figure 3.10: σ_{68} vs. bias for all the codes analyzed in Test 2. Results of Test 1 for each code are shown as a point without error bars (for simplicity) and connected to the result in this test through a solid line. This convention will also be used in subsequent plots for easier comparison against Test 1, which is the default case. All codes improve in photo-z precision (σ_{68}) due to the deeper photometry.

TEST 2: *Deep-Deep*

In this test we train or calibrate the algorithms using the Deep training sample and we apply them on the Deep testing sample (see results in Table 7). The goal of this test is to check the differences in photo-z performance when we use higher S/N data, with the caveat that this sample has also a slightly different redshift range. In order to enable an easier comparison with Test 1, we include in the plots for this Test 2 the data from the analogous plots in Test 1, and we do it by including the points for Test 1 (without error bars, for simplicity) connected by a straight line to the corresponding point for Test 2 (with error bars). Figure 3.10 shows σ_{68} versus the mean bias of the computed photo-z's for Test 2. This is analogous to Fig. 3.6. There we can see that there is general improvement in σ_{68}: most of the codes move to lower values in the plot with respect to Test 1. So we conclude that using higher S/N photometry observations increases the photometric redshift precision of a sample. However, in the case of the bias there is no general trend: some codes improve and some others do not.

Figures 3.11 and 3.12 are analogous to Figs. 3.7 and 3.8, respectively. In the plot showing the outlier fraction (Fig. 3.11), Test 2 seems to introduce an additional scattering, i.e. codes that did better in Test 1 seem to improve more in Test 2 than codes that did not do so well in Test 1, which now seem to worsen on average. This fact somehow removes the homogeneity we saw in Fig 3.7, where

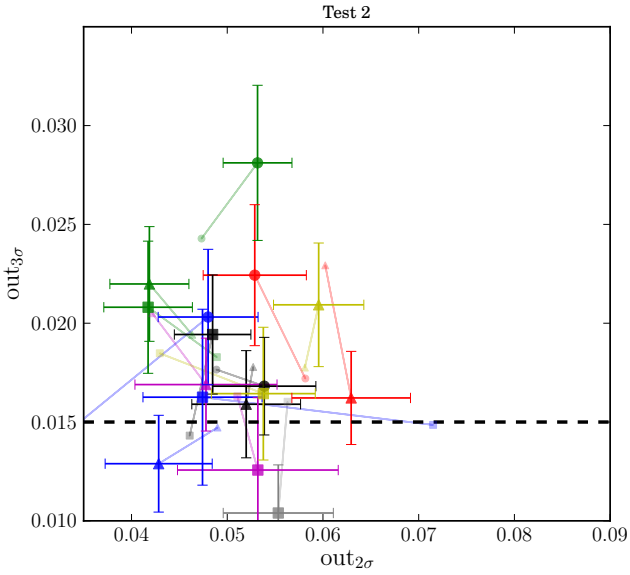

Figure 3.11: 3σ vs. 2σ outlier fraction for all the codes analyzed in Test 2. This test shows a larger scatter in the points with respect to the results in Test 1: in general, codes doing well for Test 1 improve in Test 2 while codes doing not so well for Test 1 show a worsening in Test 2.

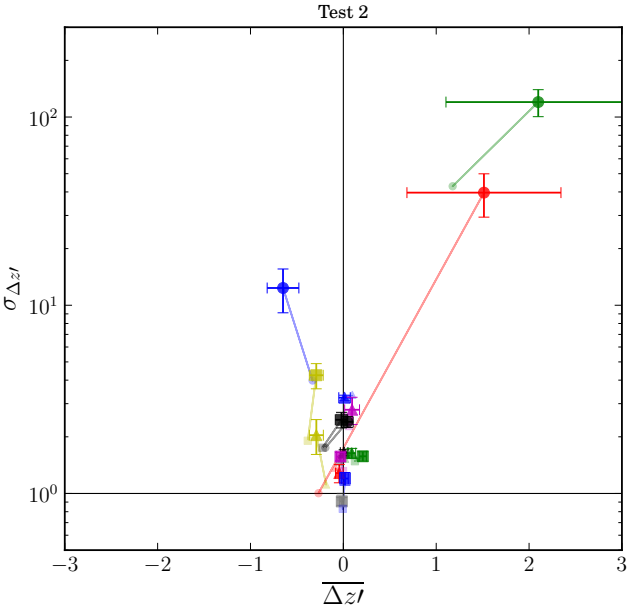

Figure 3.12: RMS vs. bias of the normalized photo-z distribution for all the codes analyzed in Test 2. Here the results are, in general, close to those observed in Test 1.

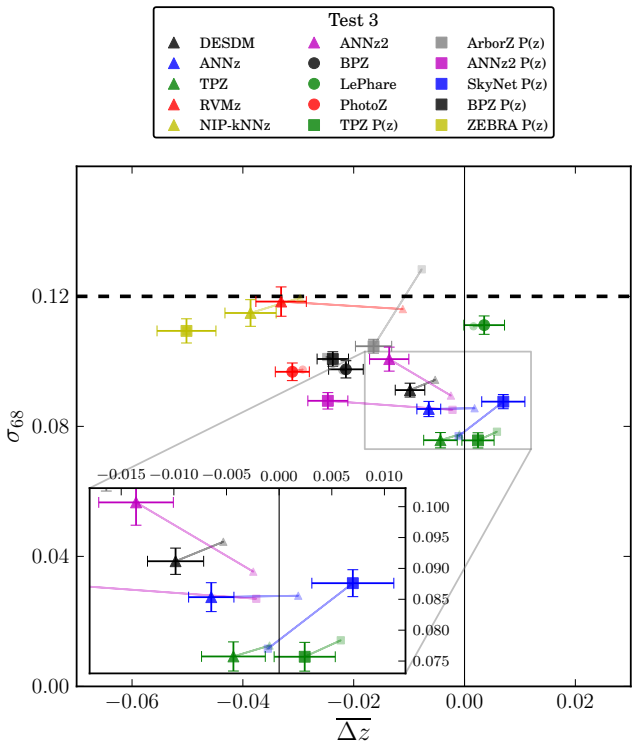

Figure 3.13: σ_{68} vs. bias for all the codes analyzed in Test 3. Results are overall compatible with Test 1 although some particular codes, such as ArborZ, show a significant improvement by using deeper data for training.

many codes produced results compatible within errors. Figure 3.12 shows the RMS and bias for the normalized Δz distribution, which, similarly to Test 1, shows that many codes underestimate the photo-z errors. The behavior here is similar to the one regarding outlier fractions: Test 2 introduces even more differences between codes doing well and codes that are less accurate.

TEST 3: *Deep-Main*

Test 3 uses the Deep training sample for training or calibration of the algorithms and uses the Main testing sample to compute photo-z's (see results in Table 8 in the Appendix). In contrast to Test 2, this is a realistic case since a Deep training sample already exists and is available to use for photo-z calibration of DES data, with the same depth as the Main testing sample. Therefore, this test explores the possibility of improving the photo-z performance of Test 1 by using higher S/N data for training or calibration.

Figure 3.13 shows the results on precision and bias for Test 3, and, as before, compares with Test 1. Some template-based codes are not included in this test so less points are shown with respect to Fig. 3.6. In this test, basically all codes fulfill the σ_{68} requirement. In addition, some codes show important improvement when using this higher S/N data for training, such as ArborZ. However, there is no general σ_{68} improvement as there was for Test 2, and the results are generally

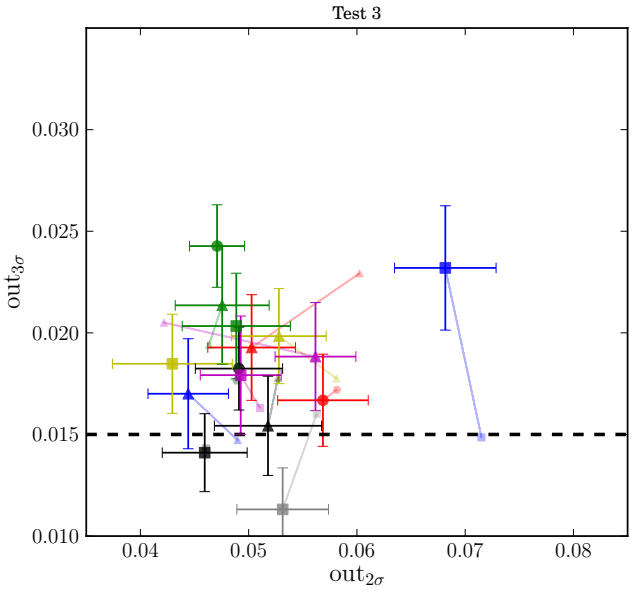

Figure 3.14: 3σ vs. 2σ outlier fraction for all the codes analyzed in Test 3. There is a general agreement with the results from Test 1.

comparable with those from Test 1. Figures 3.14 and 3.15 show also a high degree of compatibility with Test 1, contrary to what we saw in Test 2, where larger differences were appreciated.

IMPORTANCE OF THE u BAND

In this part of the chapter we want to adress the effect of adding u band photometry to the photo-z performance. It is important to stress that a u band is available in DECam, although it is not planned to be used in the DES survey.

We show the effect of incorporating the u band to Test 1. On one hand, in Fig. 3.16 we can appreciate this effect in the overall photo-z precision for the whole redshift range, and we clearly see an improvement on σ_{68} with respect to Test 1 (shown again as the points without error bars connected through a solid line) for most of the codes present in the plot.

On the other hand, since u is an ultraviolet band one expects it to be more relevant for low-redshift galaxies. Therefore, in Fig. 3.17 we show the effect of adding the u band to Test 1 in σ_{68} as a function of photometric redshift for 4 selected photo-z codes (the same ones we select for further analysis in 3.4.3). There we can clearly observe a signficant improvement in σ_{68} at low redshift ($z_{phot} < 0.5$), while the precision at higher redshift is compatible within error bars in the two cases. Furthermore, in order to visually appreciate the improvement at low redshift, in Fig. 3.18 we show the z_{phot} vs. z_{spec} scatter plot after adding the u band to Test 1 for the 4 same codes shown before, to be compared directly with Fig. 3.5.

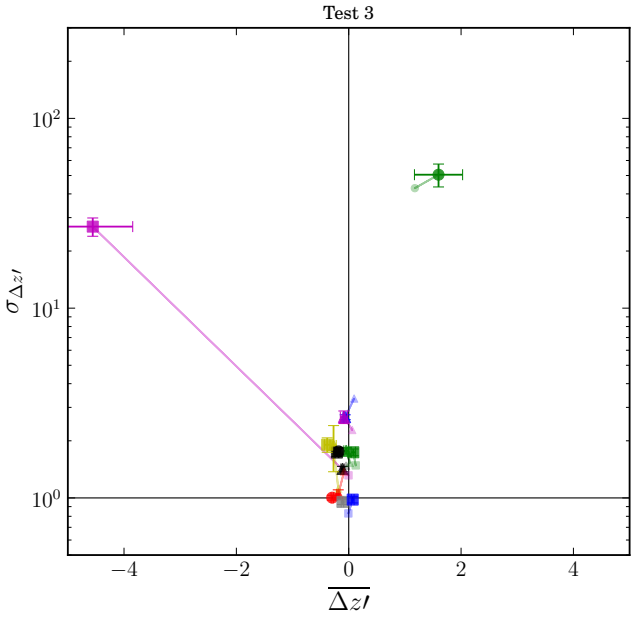

Figure 3.15: RMS vs. bias of the normalized photo-z distribution for all the codes analyzed in Test 3. Again, results generally agree with Test 1.

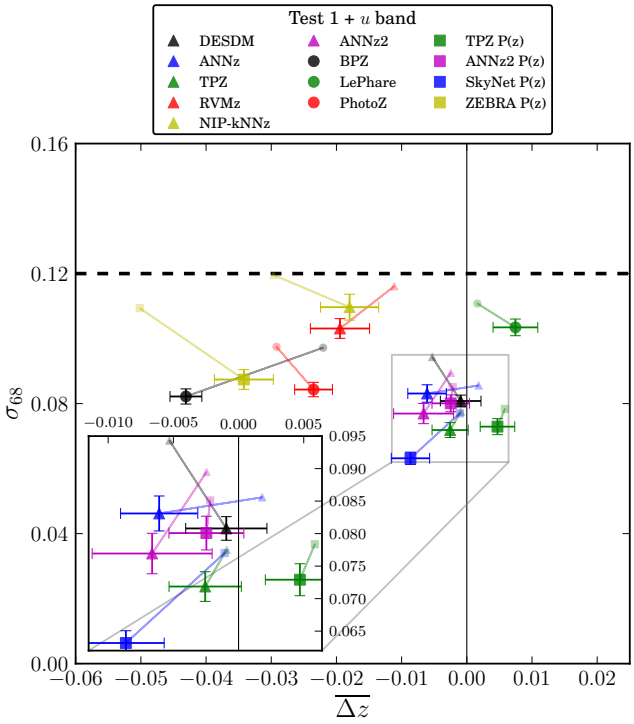

Figure 3.16: σ_{68} vs. bias for all the codes analyzed in Test 1 after the addition of the u band. We see how the incorporation of this band significantly improves the photo-z precision.

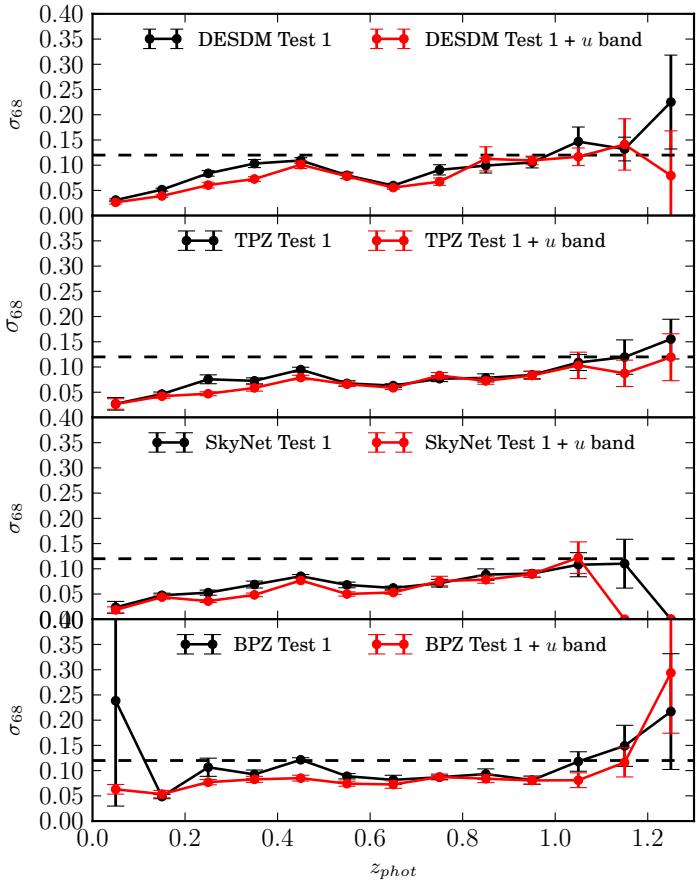

Figure 3.17: σ_{68} as a function of photometric redshift for 4 selected codes and for Test 1 + u band. As expected, the improvement in photo-z precision due to the u band addition is more important at low redshift (< 0.5).

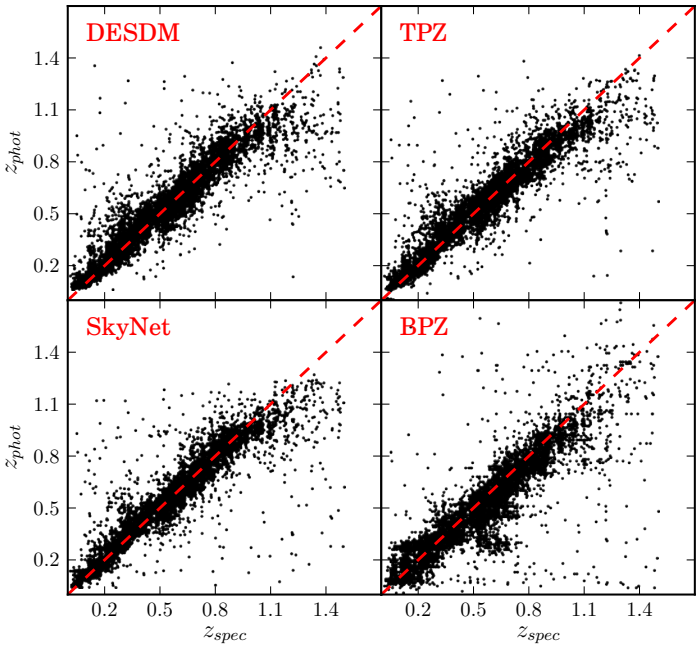

Figure 3.18: z_{phot} vs. z_{spec} scatter plot for 4 selected codes for Test 1 + u band. When comparing with the corresponding plots in Fig. 3.5, a clear improvement at low redshift, with an important reduction in the number of outliers, can be appreciated.

Table 3.4: σ_{68} for the four cases in Test 4, corresponding to training on each of the four major spectroscopic samples and testing on the full Main testing sample. The results for Test 1 are also shown, for comparison.

Codes, σ_{68}	Test 1	Test 4 VVDS Deep	Test 4 VVDS Wide	Test 4 ACES	Test 4 zCOSMOS
DESDM	0.094 ± 0.002	0.106 ± 0.002	0.139 ± 0.005	0.103 ± 0.003	0.148 ± 0.008
ANNz	0.086 ± 0.002	0.101 ± 0.003	0.138 ± 0.004	0.091 ± 0.003	0.104 ± 0.003
TPZ	0.078 ± 0.002	0.090 ± 0.002	0.110 ± 0.005	0.093 ± 0.003	0.097 ± 0.003
NIP-kNNz	0.120 ± 0.004	0.146 ± 0.006	0.156 ± 0.007	0.127 ± 0.005	0.148 ± 0.007
ANNz2	0.089 ± 0.003	0.099 ± 0.003	0.143 ± 0.008	0.104 ± 0.004	0.137 ± 0.006
BPZ	0.097 ± 0.003	0.096 ± 0.002	0.095 ± 0.002	0.095 ± 0.003	0.095 ± 0.002
LePhare	0.111 ± 0.003	0.110 ± 0.003	0.110 ± 0.003	0.111 ± 0.003	0.112 ± 0.003
PhotoZ	0.097 ± 0.003	0.101 ± 0.003	0.096 ± 0.002	0.096 ± 0.003	0.097 ± 0.003
TPZ P(z)	0.078 ± 0.002	0.091 ± 0.001	0.108 ± 0.004	0.093 ± 0.003	0.094 ± 0.002
BPZ P(z)	0.101 ± 0.002	0.096 ± 0.002	0.097 ± 0.002	0.097 ± 0.002	0.100 ± 0.002
ANNz2 P(z)	0.085 ± 0.002	0.103 ± 0.003	0.217 ± 0.009	0.103 ± 0.005	0.140 ± 0.009

TEST 4: IMPORTANCE OF DIFFERENT SPECTROSOPIC SETS

In all previous tests we have considered a single training sample (in two versions: Main and Deep) constructed with joint data from different spectroscopic surveys with different properties, as described in Section 3.3. Now we want to study the importance of the different major spectroscopic data sets used in the photo-z performance using different photo-z codes, both training and template-based. In particular, this test, that we call Test 4, consists on selecting from each of the 4 calibration fields depicted in Fig. 3.1 the major spectroscopic set used in the field, then training or calibrating the algorithms using only this subset of spectroscopic objects in the training sample, and then apply the algorithms on the full testing sample, including all the spectroscopy available. In this way, we have selected spectroscopy from VVDS Deep in the SN-X3 field, from ACES in the SN-C3 field, from VVDS Wide in the VVDS F14 field and from zCOSMOS in the COSMOS field. For simplicity, we have considered only the Main training and testing samples (i.e. main survey depth photometry), so the results should be compared with Test 1.

Table 3.4 shows the photo-z precision (σ_{68}) for the different cases in Test 4 and for Test 1, for comparison. As a first conclusion from this table, we observe how template-based codes, such as LePhare, BPZ or PhotoZ, are not very dependent on the spectroscopic data used for calibration and they show very consistent results for Test 1 and the four cases in Test 4. This is expected since these codes get the photometric information of the galaxies from a given set of predefined templates, either empirical or theoretical, and only use the galaxies in the training sample for calibration of the priors (see 3.4.1 for more information about calibration of the priors for each particular template-based code). On the contrary, and as a second

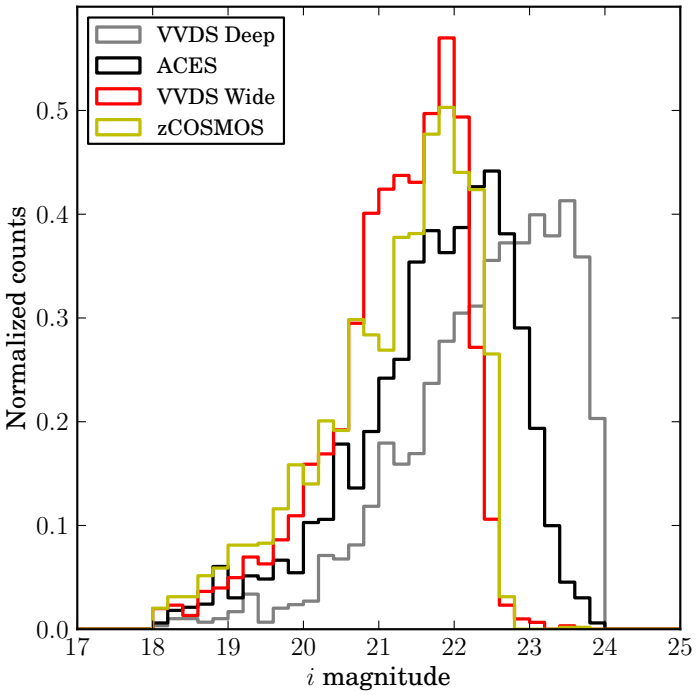

Figure 3.19: i band magnitude distributions for the four training samples used in Test 4, each corresponding only to one of the four major spectroscopic samples used, one from each of the calibration fields.

conclusion, we observe in the table how training-based codes are dependent on the data used for their training. For this class of codes (e.g. DESDM, ANNz, TPZ, ...) we can observe how the Test 1 result is generally better than any of the Test 4 results, given that, in this case, all the photometric information of the galaxies comes from the training set and thus having a more complete set helps in the photo-z performance.

Additionally, there are substantial differences in the photo-z performance depending on the spectroscopic data used for training. In order to understand this, we show in Fig. 3.19 the i magnitude distribution for each of the four training sets used in this test. On one hand, the VVDS Deep and the ACES cases of Test 4 give the most similar results to Test 1 since VVDS Deep and ACES are the deepest spectroscopic samples, and the only ones reaching $i_{AB} = 24$ as can be appreciated in Fig. 3.19. On the other hand, Test 4 VVDS Wide gives the poorest results compared to Test 1 due to the fact that the VVDS Wide spectroscopic data is much shallower ($i_{AB} < 22.5$) than DES, hence the training sample is not complete. Finally, using zCOSMOS data as a training sample seems to work better than VVDS Wide but not as well as VVDS Deep and ACES spectroscopy.

STABILITY AFTER REMOVING THE HIGHEST-WEIGHT GALAXIES

In every application of a weighting method there exists the danger of the analysis being dependent on a few, highly weighted objects in the sample. Here we

Table 3.5: σ_{68} results after removing the 5% of the galaxies with highest weights in Test 1. Also showing the default Test 1 for comparison.

Codes	Test 1	Test 1 cut
DESDM	0.094 ± 0.002	0.088 ± 0.002
ANNz	0.086 ± 0.002	0.085 ± 0.002
TPZ	0.078 ± 0.002	0.073 ± 0.002
RVMz	0.116 ± 0.004	0.103 ± 0.004
NIP-kNNz	0.120 ± 0.004	0.110 ± 0.004
ANNz2	0.089 ± 0.003	0.084 ± 0.003
BPZ	0.097 ± 0.003	0.095 ± 0.002
EAZY	0.109 ± 0.003	0.123 ± 0.002
LePhare	0.111 ± 0.003	0.109 ± 0.003
PhotoZ	0.097 ± 0.003	0.089 ± 0.003
TPZ P(z)	0.078 ± 0.002	0.073 ± 0.002
ArborZ P(z)	0.128 ± 0.003	0.117 ± 0.003
ANNz2 P(z)	0.085 ± 0.002	0.082 ± 0.002
SkyNet P(z)	0.077 ± 0.002	0.067 ± 0.002
BPZ P(z)	0.101 ± 0.002	0.097 ± 0.002
ZEBRA P(z)	0.109 ± 0.004	0.100 ± 0.004

demonstrate that the conclusions of this analysis do not change after removing the highest-weight galaxies. We test it only on the photo-z precision of Test 1, for simplicity.

In Table 3.5 we show the σ_{68} results for Test 1 together with the same test after removing the 5% of the galaxies with the highest weights (accounting for almost 30% of the total weight) in the Main testing sample. Comparing the two results for each of the photo-z codes we clearly see an improvement in photo-z precision when the 5% cut is applied, but this is expected since removing the highest-weight galaxies also means removing the faintest galaxies in the catalog, which are the most difficult galaxies to get photo-z's of. However, this predicted improvement is both small and uniform among all codes so that the main conclusions reached in the chapter remain valid.

3.4.3 RESULTS FOR DESDM, TPZ, SKYNET AND BPZ PHOTO-z CODES

So far we have compared a large number of photo-z codes in a variety of situations and configurations. Next we look in greater detail at four photo-z codes: DESDM, TPZ, SkyNet and BPZ. The DESDM photo-z code, a regression artificial neural network, is integrated within the DES Data Management service, so its results will be made available together with all the DES data products, making it

a clear choice to be studied in detail here. TPZ and SkyNet are state-of-the-art training-based methods using, respectively, random forests and artificial neural networks to compute photo-z's , and yielding the best performance among all the codes utilized in this analysis. Finally, BPZ is the template-based photo-z code showing best performance in the tests previously shown, and it has been widely used by other galaxy surveys such as CFHTLenS (Heymans et al., 2012; Hildebrandt et al., 2012). All these four codes are public.

In 3.4.2 we have studied the default configuration for this analysis, showing the most relevant quantities for all the codes and testing them against the DES requirements. However, we have not looked at the redshift dependence of these quantities, mainly due to difficulties showing that much information for a large number of codes. Figure 3.20 shows the photo-z bias, precision and outlier fractions as a function of photometric redshift for the four selected codes. We can see how the four codes behave similarly for all the metrics displayed there. The σ_{68} requirement is fulfilled by the codes in most of the redshift range, except at high redshift where error bars are large due to the small number of objects. As we previously observed, the 3σ outliers fraction is the most difficult requirement to meet, although the results are close to this limit within error bars, while the 2σ outliers fraction required is met in the whole redshift range for all the codes. As mentioned at the begining of section 3.4, the requirement on the mean bias in photo-z bins of width 0.1, $|\overline{\Delta z}| < 0.001(1+z)$, is not currently being analyzed, since it necessitates a larger spectroscopic sample in order to be able to calibrate the mean bias away. However, the top plot in Fig. 3.20 shows that for the training-based codes the mean bias in each photo-z bin is compatible with zero within errors, although the current errors are too large to assess whether the requirement is met. In some cases the overall photo-z bias is already at the 0.001 level, as can be seen in Tables 6–8, although again the errors are large.

A very important issue, which is actually the most important result needed from photo-z studies in order to perform many cosmological analyses, is the estimation of true redshift distributions $N(z)$. In Fig. 3.21 we observe how the full redshift distribution reconstructed from the four photo-z codes compares to the spectroscopic distribution. The DESDM code produces one single value for the photo-z of each galaxy in the testing sample while the other three are $P(z)$ codes, so that they return a probability density function (pdf) for each galaxy to be at a given redshift. This is the reason why the $N(z)$ reconstruction looks smoother for TPZ, SkyNet and BPZ, since these are computed from stacking all individual photo-z pdfs. Quantitatively, one can measure how good an $N(z)$ reconstruction is by looking at the $N_{poisson}$ and KS metrics in Table 6: the lower these values are, the better is the agreement between the true $N(z)$ and the photo-z-reconstructed one. As for the advantage of using $P(z)$ codes, one can observe in Table 6 how

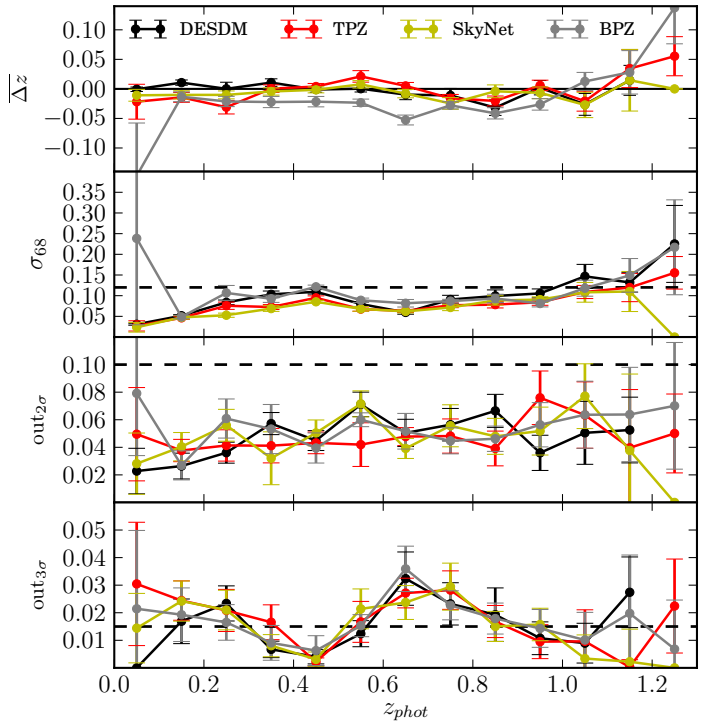

Figure 3.20: Results of Test 1 for the 4 selected photo-z codes. From top to bottom, the photo-z bias, σ_{68} and the 2σ and 3σ outlier fractions, in bins of z_{phot}. DES requirements, displayed as a black dashed line, are shown for the latter 3 metrics.

the $\mathrm{N}_{\mathrm{poisson}}$ values for TPZ and BPZ are significantly smaller in their $P(z)$ versions than in their single-estimate photo-z versions.

On the other hand, although this full redshift distribution is interesting for photo-z analyses, most of the cosmological studies split the galaxy sample into multiple photo-z bins, therefore there is a need to know the true redshift distribution inside each of those photo-z bins. Figure 3.22 shows the redshift distributions, both spectroscopic and photometric, for six photo-z bins of width 0.2 from $z = 0.1$ to $z = 1.3$, and for the four photo-z codes selected. The limited number of spectroscopic galaxies available makes the distributions shown in the figure somewhat noisy, especially in the last photo-z bin, where a very small number of galaxies is available. The third and fourth bins in photo-z are the ones presenting the narrowest spectroscopic redshift distributions, which agrees with the fact that the photo-z precision is the highest in this redshift range as can be appreciated in Fig. 3.20.

In Fig. 3.22, we observe how single-estimate photo-z codes produce a top-hat photo-z distribution for each (photo-z selected) redshift bin. In this case, depicted in the left column of Fig. 3.22, the photometric and spectroscopic redshift distributions of each bin are very different and therefore a spectroscopic sample is needed to calibrate the broadening of the redshift bin due to photo-z errors. On the other hand, when using $P(z)$ codes to bin a sample in photometric redshift, one selects a galaxy to be inside a given redshift bin by looking at the position of

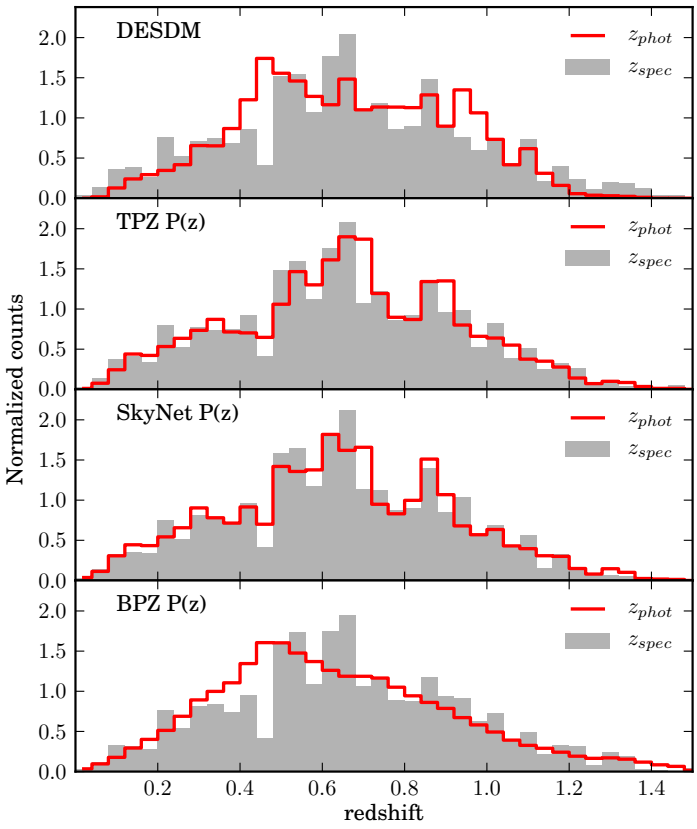

Figure 3.21: Full weighted spectroscopic redshift distribution and its photo-z reconstruction using the four selected codes for Test 1. TPZ, SkyNet and BPZ produce redshift pdfs for each galaxy, thus yielding smoother photo-z distributions.

the median of the pdf (other choices are also possible, e.g. the mode), checking whether it is within the boundaries of the the bin and summing the full pdf of the galaxies inside, including probabilities beyond the bin limits. That makes the photo-z distribution broader than the bin limits and closer to the spectroscopic redshift distribution of the bin, as can be seen in the three rightmost columns in Fig. 3.22. We can see on those panels how the tails of the spectroscopic distributions are well represented by the photo-z distributions. This is an important point in favor of $P(z)$ codes since their ability to reproduce the spectroscopic redshift distribution of a photo-z selected bin by stacking their redshift pdfs makes them less dependent on a spectroscopic calibration sample.

In summary, we have characterized for each code the true redshift distribution inside each photo-z bin. This is the most important quantity to be extracted from any photo-z analysis, since it is the starting point for many cosmological studies such as galaxy clustering and weak lensing. Regarding the performance in such task, the four codes studied in this section show similar spectroscopic redshift distributions for each photo-z bin, but $P(z)$ codes are able to yield a better reconstruction of these distributions by adding up the redshift pdfs for each galaxy which makes them somewhat less reliant in the precise photo-z calibration.

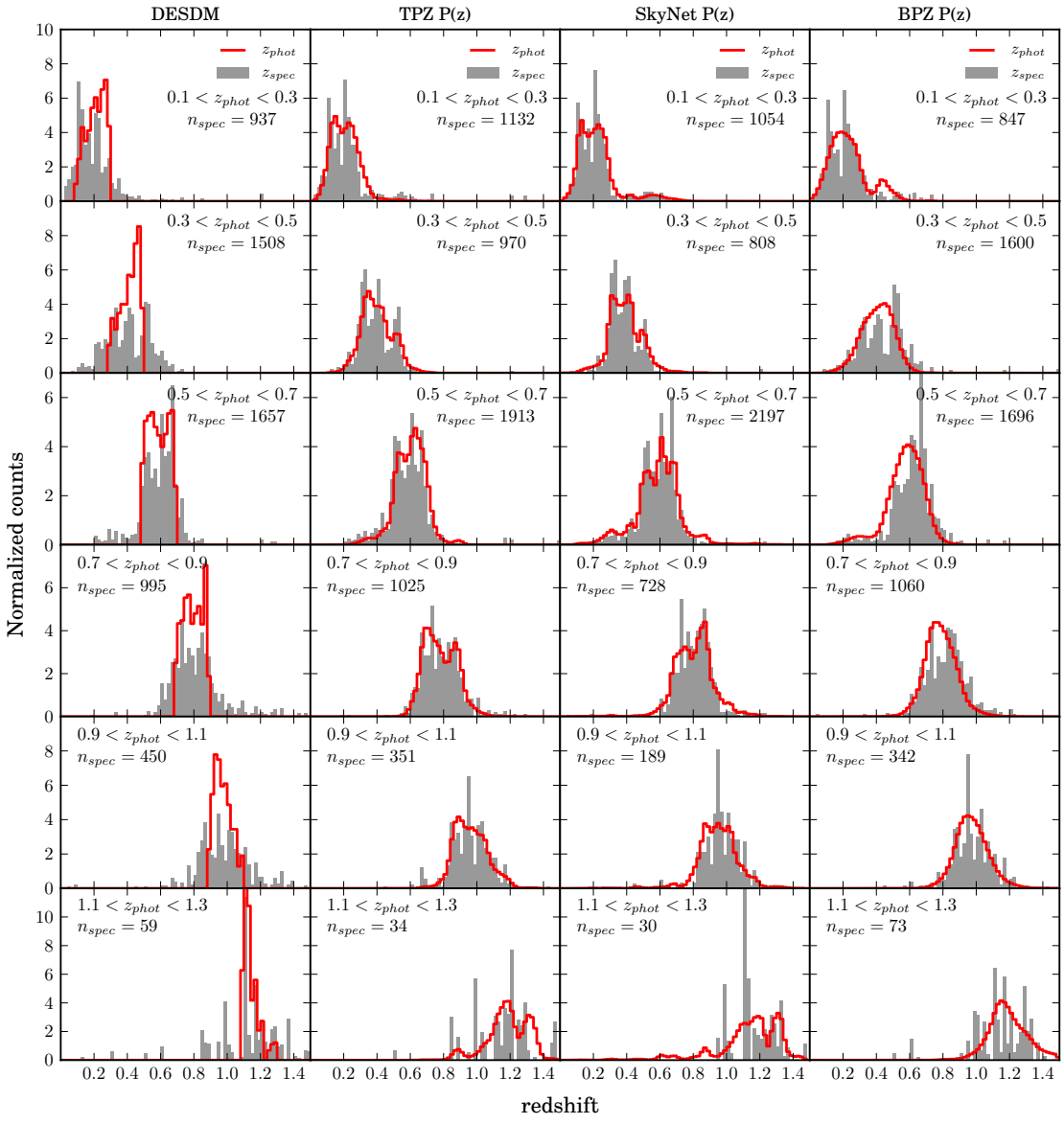

Figure 3.22: Weighted spectroscopic redshift distributions and their photo-z reconstruction using the four selected codes, for photo-z bins of width 0.2. The number of spectroscopic galaxies inside each photo-z bin is shown. The DESDM is a single-estimate photo-z code, while TPZ, SkyNet and BPZ are $P(z)$ codes. This is the reason why the photo-z distributions returned by the latter codes can reconstruct the tails of the spectroscopic distributions beyond the photo-z bins. The photo-z bins are defined using the best estimate z_{phot} for each code, while, for TPZ, SkyNet and BPZ the recostructed redshift distributions are obtained by stacking the probability density functions for each galaxy.

3.5 DISCUSSION

We now discuss some of the results and implications of the analyses presented above. First, we consider the results from different types of photo-z codes, and afterwards we compare the main outcomes from this study with previous results in the literature.

The photo-z codes showing the best performance in the analysis are all training-based methods. Among them, there are various codes using Artificial Neural Networks (ANNs) in different ways and configurations (see Section 3.4.1), and the similarities and differences between them go beyond the network architecture. DESDM uses hyperbolic tangent activation functions in the hidden layer while ANNz and SkyNet use a sigmoid activation function. ANNz2 results are a mean of several runs where both activations functions are used in the different runs. Vanzella et al. (2004) came to the conclusion that the choice of activation between sigmoid and hyperbolic tangent functions has little effect on the photo-z performance. SkyNet was also run with rectified linear units as activation function, which have been shown to outperform more traditional activation functions in object classification (Glorot, Bordes & Bengio, 2011), but no improvement in the photo-z problem was observed. SkyNet, ANNz and ANNz2 use weight regularization to avoid over fitting while all the four methods monitor the performance on a validation set to prevent overtraining. DESDM and ANNz use first-order gradient information while SkyNet and ANNz2 also use second-order derivative information to train the network. SkyNet and ANNz2 are the only networks adding a constraint to the outputs: ANNz2 does this on a redshift bin per redshift bin basis while SkyNet uses a *softmax* transformation in the final layer, adding a constraint on all redshift bins simultaneously. In conclusion, the neural networks with regularization perform better than the un-regularized DESDM network while the fact that SkyNet uses a *softmax* constrained output in combination with a back-propagation algorithm that uses second order derivative information seems to give it the edge over the rest of ANNs.

Aside from ANNs, TPZ, which is a state-of-the-art photo-z code using Prediction Trees and Random Forests, performs remarkably well in all the tests in this work. The prediction trees and random forest techniques used by TPZ have the advantage that they have fewer hyper parameters to be chosen compared to neural networks. Neural networks have, amongst others, to choose the amount of hidden layers, the amount of nodes per hidden layer, the learning rate and at least one regularization parameter if present. Random forests used in TPZ have only 2 hyper parameters to choose: the amount of trees used and the size of the subsample set of features used at each split. This leaves out the choice of activation function in neural networks and the choice of the measure of information

gain at each split in random forests, maximizing its performace.

Furthermore, training-based photo-z codes show lower bias compared to that of template-based codes, which indicates possible systematic inaccuracies in the template sets. This can be solved by using adaptive recalibration procedures, which adjust the zero-point offsets in each band using the training sample. Such technique has been succesfully applied by LePhare in this work, as was also the case in Hildebrandt et al. (2010).

The photo-z precision values obtained by template-based methods in this study are all compatible with each other within 10%. BPZ and PhotoZ yield the highest photo-z precision among the methods of this class. The reason for this is probably not the template sets they use, which for BPZ is a combination of Coleman, Wu & Weedman (1980) and Kinney et al. (1996), and a more complex combination of templates for PhotoZ (see 3.4.1), since other codes include similar libraries. The fact that they both use Bayesian priors calibrated on DES data, and not only on previous datasets, can be relevant here, although this makes the result more dependent on how representative the training sample is of the full DES data sample.

In order to set up a context for DES photo-z results, it is worth checking the performance obtained by previous similar surveys. Of particular interest is the comparison to the CFHTLS and CFHTLenS (which are two different reductions of the same survey), photo-z results (Coupon et al., 2009; Hildebrandt et al., 2012), due to the similarities between CFHTLS and DES in terms of survey specifications. However, it is important to stress the differences between DES and CFHTLS. For instance, CFHTLS uses deep u band photometry by default, although they also test the importance of this band by computing their photo-z's without it (Ilbert et al., 2006). They find a clear degradation in their photo-z precision at $z_{spec} < 0.4$, compatible with our findings. Another difference comes from the fact that they do not apply any weighting technique to their calibration sample, thus leaving room for discrepancies between their calibration and full samples. In addition we should point out that their spectroscopic samples for training/calibration are much larger than the ones used here by about a factor of 5. Finally, their photometry (S/N$\simeq$ 11 at $i_{AB} = 24$) is deeper than the Main reduction in this chapter, being close to our Deep sample (see Fig. 3.2). In Coupon et al. (2009), CFHTLS photo-z's show a precision of about $\sigma_{68} \simeq 0.085$ (roughly translating their result to the metric used here) for $ugriz$ photometry in galaxies up to $i_{AB} < 24$. So, despite the differences stated above, their value of the photo-z precision is at the level of the DES photo-z precision shown in this chapter. Moreover, this level of precision is obtained already in Test1, which uses slightly shallower photometry and lacks the u band. Even so, the differences in photo-z precision obtained by different algorithms might be relevant here. In particular, the results reported in Coupon et al. (2009) used a template-based photo-z code, while the most pre-

cise methods reported in this study come from training-based codes. If we do the comparison between template-based codes, the results in this study are close but do not exceed the precision reported by CFHTLS.

In terms of outlier fraction, Coupon et al. (2009) report a value for the fraction of galaxies with $|\Delta z/(1+z)| > 0.15$ of $\eta = 10.1\%$. In order to enable the comparison, since the metric used in this chapter differs from their approach, we have computed η for two cases in this study. On the one hand, we have estimated η for a template-based photo-z code (BPZ), trying to make a fair comparison since this is the class of algorithm used in Coupon et al. (2009). In this case, we obtain $\eta = 10.0\%$, in perfect agreement with their result. On the other hand, computing this outlier rate for a template-based code (DESDM) yields $\eta = 6.1\%$, significantly improving the template-based code result.

We conclude that the overall photo-z results shown in this chapter are within the expectations for DES if we take into account the performance reported by previous comparable studies for a survey of similar characteristics.

3.6 SUMMARY AND CONCLUSIONS

The Dark Energy Survey succesfully installed DECam during the second semester of 2012, starting its operations in November 2012 with a Science Verification (SV) period lasting until February 2013. Among the 150 sq. deg. covered by SV observations, four different fields, of about 3 sq. deg. each, overlap with areas with substantial spectroscopic coverage such as VVDS 02hr or COSMOS. Using $grizY$ photometry for galaxies matched to the existing spectroscopic data in these four calibration fields, this chapter presents the photometric redshift performance of the DES survey in the SV period. Most of the relevant photo-z codes have been used in the analysis.

Since spectroscopic galaxy samples are generally shallower, a weighting technique is used to make the calibration sample of galaxies to mimic the DES full sample in magnitude and color space in order to properly estimate the photo-z performance in the DES galaxy sample.

Calibration and testing samples have been produced with two different depths: Main is the default depth in the DES survey, and Deep corresponds to the depth in SNe fields. Test 1, which uses the Main training and testing samples, represents the default case for photo-z estimation in DES. Results from 13 different codes are analyzed in this case, showing fluctuations in photo-z performance but a general agreement in codes of the same type (machine-learning or template fitting algorithms). In particular, most of the codes analyzed comfortably meet the DES science requirements in terms of photo-z precision and several also meet the

requirements on the fractions of outliers.

In Test 2 we explore the impact of deeper and higher S/N photometry in photo-z calculations, showing that all the codes used improve their results significantly, as expected. In Test 3 we explore the possibilty of using deeper photometry only for training/calibration of the algorithms. In this case we see no general improvement, although there is a significant enhancement using specific codes.

In an additional test, we consider the incorporation of the u band, which is available in DECam but not used in the DES survey, demonstrating a general improvement in photo-z precision, particularly at low redshift (< 0.5). This is expected since the u band is crucial for the filter set to bracket the 4000Å break in low redshift galaxies. However, due to the high mean redshift of the full DES sample, the impact of not using that band in the overall photo-z precision is less important than in previous, shallower surveys such as SDSS. Moreover, we study the importance of the different spectroscopic data sets used, showing how the sets spanning the whole photometric space are crucial for training-based methods, and demonstrated that the results are stable under the removal of the galaxies with highest weights from the analysis.

Generally, training-based photo-z codes show the best performance in the tests in terms of photo-z precision and bias. Among them, TPZ, using Prediction Trees and Random Forest, and SkyNet, a state-of-the-art Artificial Neural Network application, seem to yield the most accurate results, achieving a core photometric redshift resolution below $\sigma_{68} = 0.08$. The fact that these two new codes perform better than others extensively used in the literature shows how there is room for improvement in the photo-z industry. On the other hand, all template fitting methods employed show consistent results between them, although the use of Bayesian priors specifically calibrated on DES data and adaptive template-recalibration procedures appears to help significantly.

Finally, in the last part of the chapter, we choose four photo-z codes, representing different techniques and types, and we present a more detailed analysis of their results. We show some of the most relevant metrics as a function of redshift and, most importantly, we study the estimation of the true redshift distributions $N(z)$ computed using photo-z's . For these four codes, we obtain the true redshift distributions in six photo-z bins. Figure 3.21 shows the DES photo-z capabilities in such crucial task, and demonstrates the ability to split the DES full sample in tomographic redshift bins of width 0.2 already with these early data. Furthermore, the calibration of the true redshift distribution of a photo-z selection is the most important ingredient for cosmological studies involving galaxy clustering or weak lensing, and it is an important outcome from this chapter, enabling further science analyses.

The photo-z analyses carried out in this work using these early stage DES data

will serve as a benchmark for future data releases, and as the survey area grows during the observation period, more spectroscopic data will be available allowing a better calibration and a better sampling for training algorithms. Therefore these promising early results will do nothing but improve in the near future, which will allow putting tighter constrains on several cosmological parameters. Furthermore, the 5-band optical and near-infrared photometry of DES can be combined with the infrared J and K_s photometry provided by the VHS survey (McMahon et al., 2013) in 90% of the DES footprint. This should result in improved photometric redshift estimations, particularly at high redshift (Banerji et al., 2008).

Part III

Cosmology from LSS and WL

Chapter 4

COSMOLOGY FROM LARGE SCALE GALAXY CLUSTERING AND GALAXY-GALAXY LENSING WITH DES-SV DATA

4.1 INTRODUCTION

Since the discovery of cosmic acceleration, the nature of dark energy has emerged as one of the most important open problems in cosmology. Wide-field, large-volume galaxy surveys are promising avenues to answer cosmological questions, since they provide multiple probes of cosmology, such as Baryon Acoustic Oscillations (BAO), large scale structure, weak lensing and cluster counts from a single dataset. Moreover, some of these probes can be combined for greater effect, since each is sensitive to their own combination of cosmological parameters and systematic effects. In this chapter, we will focus on combining the large scale angular clustering of galaxies with measurements of the gravitational lensing produced by the large scale structure traced by the same galaxies, as observed in the Dark Energy Survey (DES).

Measurements of the large scale clustering of galaxies are among the most mature probes of cosmology. The positions of galaxies are seeded by the distribution of dark matter on large scales and the manner in which the growth of structure proceeds from gravitational collapse is sensitive to the relative amounts of dark matter and energy in the Universe. There is a long history of using large-volume galaxy surveys for the purposes of constraining cosmology, including DES, Sloan Digital Sky Survey (SDSS) (York et al. , 2000), Hyper Suprime-Cam (HSC) (Miyazaki et al., , 2012), the Kilo-Degree Survey (KiDS) (de Jong et al., 2013; de Jong et al. , 2015; Kuijken et al., 2015), and the Canada France Hawaii Telescope Lensing Survey (CFHTLenS) (Heymans et al. , 2012; Erben et al., 2013).

Gravitational lensing, the deflection of light rays by massive structures, provides a complementary method of probing the matter distribution. Here we focus on galaxy-galaxy lensing (Tyson et al., 1984; Brainerd, Blandford & Smail , 1996), when both the lenses and sources are galaxies. This involves correlating the amount of distortion in the shapes of background galaxies with the positions of foreground galaxies. The amount of distortion is indicative of the strength of the gravitational potential along the line of sight and therefore tells us about the amount of matter contained in the lens plane. Weak gravitational lensing produces two effects, magnification of the source and shearing of its image, but this analysis is only concerned with the latter. These have been used to probe both cosmology (Mandelbaum et al., 2013; Cacciato et al., 2013; More et al., 2015) and the structure of dark matter halos and its connection to the galaxy distribution and baryon content of the Universe (Sheldon et al., 2004; Mandelbaum et al., 2006, 2008; Cacciato et al., 2009; Leauthaud et al., 2012; Gillis et al., 2013; Velander et al., 2014; Hudson et al., 2015; Sifón et al., 2015; Viola et al., 2015; van Uitert et al., 2016).

Individual studies of large scale structure (Crocce et al., 2015), galaxy-galaxy lensing (Clampitt et al. , 2016) and cosmic shear (Becker et al., 2015; The Dark Energy Survey Collaboration , 2015) using DES data as well as combined analyses focusing on smaller scales (Park et al., 2015) have been presented elsewhere. In this chapter, we combine angular clustering and galaxy-galaxy lensing to jointly estimate the large-scale galaxy bias and matter clustering and constrain cosmological parameters.

The plan of the chapter is as follows. Section 4.2 outlines the theoretical framework for modelling the angular galaxy correlation function and galaxy-galaxy lensing. Section 4.3 describes the galaxy sample used and the measurements from DES data, as well as the covariance between the two probes. Our cosmology results are summarized in Section 4.4 including constraints on a five-parameter ΛCDM (Cold Dark Matter) model and a six-parameter wCDM model, where w, the dark energy equation of state parameter is also allowed to vary. We discuss the robustness of our results and our tests for systematic errors in Section 4.5. Finally, we combine our analysis with other probes of cosmology and compare our results to previous results in the literature in Section 4.6. Our conclusions are presented in Section 4.7.

4.2 THEORY

We are interested in describing the angular clustering of galaxies, $w(\theta)$, and the tangential shear produced by their host dark matter halos, $\gamma_t(\theta)$, as a function

of cosmology. The angular correlation function, $w(\theta)$, can be expressed in terms of the galaxy power spectrum as:

$$C(\ell) = \frac{1}{c}\int d\chi \left(\frac{n_l(\chi)H(\chi)}{\chi}\right)^2 P_{gg}(\ell/\chi), \tag{4.1}$$

$$w(\theta) = \int \frac{\ell d\ell}{2\pi} C(\ell) J_0(\ell\theta), \tag{4.2}$$

where P_{gg} is the galaxy auto power spectrum, J_0 is the Bessel function of order 0, l is the angular wavenumber, χ is the comoving radial co-ordinate, $H(\chi)$ is the Hubble relation, c is the speed of light, and $n_l(\chi)$ is the number of galaxies as a function of radial distance from the observer, normalized such that $\int_{\chi_{\rm min}}^{\chi_{\rm max}} n_l(\chi)\, d\chi = 1$. Note that Eq. 4.2 uses the Limber approximation (Limber , 1953; Kaiser , 1992), such that the radial distribution of galaxies, $n_l(\chi)$, is assumed to be slowly varying over our redshift slice. We have also ignored the contribution of redshift-space distortions to the angular clustering; this is expected to be small due to the width of the redshift intervals used; for the full expression, see Crocce et al. (2015).

The tangential shear is given by:

$$\langle\gamma_t(\theta)\rangle = 6\pi\Omega_m \int d\chi\, n_l(\chi)\frac{f(\chi)}{a(\chi)}\int dk\, k P_{g\delta}(k,\chi) J_2(k,\theta,\chi), \tag{4.3}$$

where $f(\chi) = \int d\chi' n_s(\chi')\chi(\chi-\chi')/\chi'$ is the lens efficiency, a is the scale factor and $n_l(\chi)$ and $n_s(\chi)$ are the selection functions of the lenses (foreground) and source (background) galaxies respectively. The foreground galaxies supply the gravitational potentials that lens the background galaxies. The tangential shear is a measurement of the amount of distortion introduced into the images of background galaxies from the gravitational potentials along the line of sight as a function of scale. We will discuss the impact of photometric redshift (photo-z) errors on the lens and source distributions and propagate these to the measured cosmological constraints in Section 4.5.

The combination of these two probes has been extensively discussed in the literature (Baldauf et al., 2010; Yoo & Seljak, 2012; Mandelbaum et al., 2013; Park et al., 2015) and provide another means by which we can mine the rich, well calibrated DES-SV dataset. Unlike Park et al. (2015), we restrict our modelling to sufficiently large scales such that we are not sensitive to how galaxies populate individual halos, i.e. Halo Occupation Distribution (HOD) modelling is unnecessary. On these scales, we are only concerned with correlations between galaxies that reside in different halos (the 2-halo term of the power spectrum), and we can relate the matter power spectrum, $P_{\delta\delta}$, to the galaxy power spectrum, P_{gg}, and

galaxy-dark matter cross-power spectrum, $P_{g\delta}$, via the following relationships:

$$P_{gg}(k) \approx b_g^2 P_{\delta\delta}(k), \tag{4.4}$$

$$P_{g\delta}(k) \approx b_g r P_{\delta\delta}(k), \tag{4.5}$$

where b_g is the linear bias that relates the clustering of galaxies to that of dark matter and r is the cross-correlation coefficient that captures the stochasticity between the clustering of dark matter and the clustering of galaxies; see for example Seljak (2000); Guzik & Seljak (2001).

The measurement of $w(\theta)$ depends on $b_g^2 P_{\delta\delta}$, while the tangential shear, $\gamma_t(\theta)$, depends on $b_g P_{\delta\delta}$ if $r = 1$, a reasonable approximation on the large scales we use in this work (we allow for and marginalize over possible stochasticity through our non-linear bias modelling; see Section 4.2.1). The measurements of $w(\theta)$ and $\gamma_t(\theta)$ in combination allow us to estimate both the clustering amplitude and the linear galaxy bias, thus enabling us to obtain useful cosmological information.

4.2.1 NON-LINEAR BIAS MODEL

The assumption of linear bias in Eqs. (4.4) and (4.5) is expected to break down at small scales. In order to account for this effect, we use the non-linear biasing scheme of McDonald (2006), where the galaxy over-density, δ_g, is written as

$$\delta_g = \epsilon + b_1\delta + b_2\delta^2 + \text{next leading order bias terms}, \tag{4.6}$$

where b_1 is the usual linear bias, b_2 is the next leading order bias term and ϵ is the shot noise. The bias parameters, b_1 and b_2 are not known a priori and become free parameters to be constrained during the analysis. Under this perturbation theory scheme, the galaxy-dark matter and galaxy-galaxy power spectra become

$$P_{g\delta} = b_1 P_{\delta\delta} + b_2 A(k), \tag{4.7}$$

$$P_{gg} = b_1^2 P_{\delta\delta} + b_1 b_2 A(k) + b_2 B(k) + N, \tag{4.8}$$

where N is the shot noise and $A(k)$ and $B(k)$ can be calculated using standard perturbation theory as follows:

$$A(k) = \int \frac{d^3q}{(2\pi)^3} F_2(\mathbf{q}, \mathbf{k}-\mathbf{q}) P_{\delta\delta}(q) P_{\delta\delta}(|\mathbf{k}-\mathbf{q}|), \tag{4.9}$$

$$B(k) = \int \frac{d^3q}{(2\pi)^3} P_{\delta\delta}(q) P_{\delta\delta}(|\mathbf{k}-\mathbf{q}|), \tag{4.10}$$

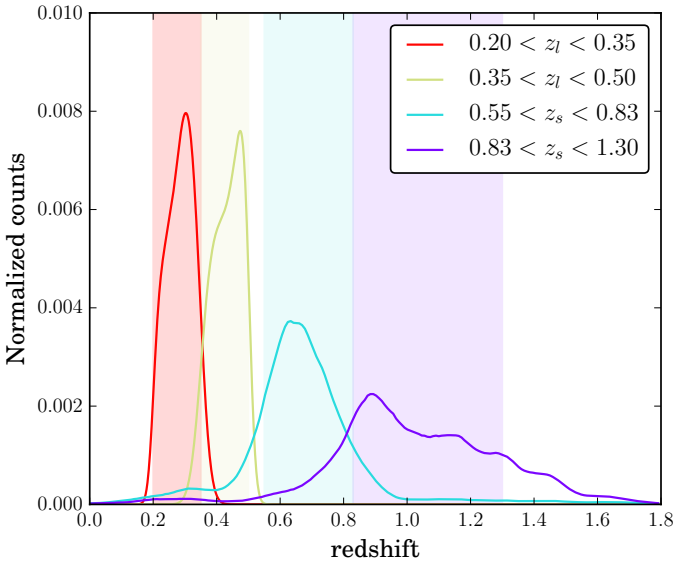

Figure 4.1: Redshift distributions of the four galaxy samples used in this work. Red and yellow curves correspond to the two redMaGiC lens bins while cyan and purple curves correspond to the two source bins in the fiducial configuration (`ngmix` shears, SkyNet photo-z's).

where $F_2(\mathbf{k_1},\mathbf{k_2}) = \frac{5}{7}\frac{(\mathbf{k_1}+\mathbf{k_2})\cdot\mathbf{k_2}}{k_1^2} + \frac{1}{7}(\mathbf{k_1}+\mathbf{k_2})^2\frac{\mathbf{k_1}\cdot\mathbf{k_2}}{k_1^2 k_2^2}$. Note that this non-linear biasing scheme generates departures from $r = 1$ as $r \approx 1 - 1/4(b_2/b_1)^2\xi_{gg}$, where ξ_{gg} is the correlation function. As such we do not include an additional free parameter for the cross-correlation coefficient. We found that for reasonable values of the shot noise, N, given the density of our galaxy sample, has a less than 5% effect on $w(\theta)$ on scales below our regime of interest ($< 20'$) and so have ignored this term for the remainder of our analysis. We do, however, include an additional additive constant term in configuration space as discussed in Section 4.5. This term mainly alters the large scale clustering to allow for possible systematics coming from observational effects (see Section 4.5.6).

We investigate the inclusion of the next order biasing term in Section 4.5.1, in which we vary both the lower limit on the angular scale cutoff and the modelling of non-linear bias.

4.3 DATA AND MEASUREMENTS

The Dark Energy Survey (DES) is an ongoing photometric survey that aims to cover 5000 sq. deg. of the southern sky in five photometric filters, $grizY$, to a depth of $i \sim 24$ over a five year observational program using the Dark Energy Camera (DECam, Flaugher et al. (2015)) on the 4m Blanco Telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile. In this analysis, we will be utilizing DES-SV (Science Verification) data, in particular a contiguous $\sim$ 139 sq. deg. patch known as the SPT-E region (because of its overlap with the South Pole Telescope survey footprint). This is only a small ($\sim$ 3%) subset of the

expected eventual sky coverage of DES, but observations in all five filters have been performed at full depth, although substantial depth variations are present (see e.g. Leistedt et al. 2015), mainly due to weather and early DECam operational challenges. The DES-SV data have been used for constraining cosmology in this work, but a rich variety of science cases are possible with this data sample (see The Dark Energy Survey Collaboration (2016) and references therein).

The lens galaxy sample used in this work is a subset of the DES-SV galaxies selected by redMaGiC[1] (Rozo et al., 2015b), which is an algorithm designed to define a sample of Luminous Red Galaxies (LRGs) by minimizing the photo-z uncertainty associated with the sample. It selects galaxies based on how well they fit a red sequence template, as described by their goodness-of-fit, χ^2. The red sequence template is calibrated using redMaPPer (Rykoff et al., 2014; Rozo et al., 2015a) and a subset of galaxies with spectroscopically verified redshifts. The cut-off in the goodness of fit, $\chi^2_{\rm cut}$, is imposed as a function of redshift and adjusted such that a constant comoving number density of galaxies is maintained, since red galaxies are expected to be passively evolving. The redMaGiC photo-z's show excellent performance, with a median photo-z bias, $(z_{\rm spec} - z_{\rm phot})$, of 0.005 and scatter, $\sigma_z/(1+z)$, of 0.017. Equally important, their errors are very well characterized, enabling the redshift distribution of a sample, $N(z)$, to be determined by stacking each galaxy's Gaussian redshift probability distribution function (see Rozo et al. 2015b for more details).

The galaxy shape catalogs used in this work were presented in Jarvis et al. (2015), and they have been used in several previous analyses (Vikram et al., 2015; Becker et al., 2015; The Dark Energy Survey Collaboration , 2015; Gruen et al., 2015; Clampitt et al. , 2016). Two different catalogs exist corresponding to the `ngmix` [2] (Sheldon, 2014) and `im3shape` [3] (Zuntz et al., 2013) shear pipelines, both producing model fitting shape measurements to a subset of DES-SV galaxies. The two catalogs differ in their approach to modelling the intrinsic galaxy shape (`ngmix` uses a Gaussian mixture model to approximate an exponential disk galaxy profile while IM3SHAPE determines the maximum likelihood for fitting a bulge and/or disk profile) and also in the number of filters used (`ngmix` uses riz bands while IM3SHAPE only uses r band). This results in the `ngmix` catalog containing more sources than IM3SHAPE ($\sim$6.9 galaxies per arcmin^2 vs. $\sim$4.2 galaxies per arcmin^2). More details about the pipelines and an extensive set of null and systematics tests can be found in Jarvis et al. (2015). The photo-z distributions of the galaxies in the shear catalogs were studied in detail in Bonnett et al. (2015), using 4 different photo-z codes that performed well in a previous more extensive

[1] `https://des.ncsa.illinois.edu/releases/sva1`
[2] `https://github.com/esheldon/ngmix`
[3] `https://bitbucket.org/joezuntz/im3shape`

photo-z code comparison (Sánchez et al., 2014). The four methods are SkyNet (Graff et al., 2014; Bonnett et al., 2015), ANNz2 (Sadeh et al. , 2015), TPZ (Carrasco Kind & Brunner , 2013) and BPZ (Benítez, 2000). The first three methods are training-based, and the last is a widely used template-based code. Details about their training or calibration procedures and about the validation against spectroscopic data can be found in Bonnett et al. (2015).

In this chapter we use the `ngmix` shear catalog and SkyNet photo-z's for the fiducial results, but we will test the robustness of our results with the IM3SHAPE shear catalog as well as using the source distributions derived from the other photo-z algorithms in the analysis.

4.3.1 MEASUREMENTS

We use two lens bins, selected using redMaGiC photo-z's: $0.20 < z < 0.35$ and $0.35 < z < 0.50$, and two source bins, selected using SkyNet photo-z's: $0.55 < z < 0.83$ and $0.83 < z < 1.30$. The same lens photo-z bins are analyzed in Clampitt et al. (2016) while the source photo-z bins are studied in detail in Bonnett et al. (2015) and used for cosmology in The Dark Energy Survey Collaboration (2015). Individual analyses involving $\gamma_t(\theta)$ and $w(\theta)$ with DES-SV have been presented in Clampitt et al. (2016) and Crocce et al. (2015), respectively. Figure 4.1 shows the redshift distributions for the lens and source bins utilized in this analysis. For each lens bin, we measure the galaxy clustering and the galaxy-galaxy lensing signals using the estimators defined next. The correlation functions have been estimated using the code `TreeCorr`[4] (Jarvis et al. , 2004).

ANGULAR CLUSTERING – $w(\theta)$

On the galaxy clustering side, we compute the angular correlation function for each redshift bin using the minimum variance estimator of Landy & Szalay (1993),

$$w(\theta) = \frac{\mathrm{DD} - 2\mathrm{DR} + \mathrm{RR}}{\mathrm{RR}}, \tag{4.11}$$

where θ is the angular separation in the sky, and DD, DR and RR are data-data, data-random and random-random pairs of galaxies, with data and random galaxies having the exact same geometry in the sky. For the random catalogs, we use 136185 random points (~ 7 and ~ 15 times the number of galaxies in the fiducial bin and low-z bin respectively) and apply the same angular masking as the redMaGiC galaxies in the SV region. The resulting measurement is shown in Fig. 4.2. The clustering amplitude falls from $\sim 10^{-1}$ to 10^{-2} over the range $\theta = 10 - 100$

[4] `https://github.com/rmjarvis/TreeCorr`

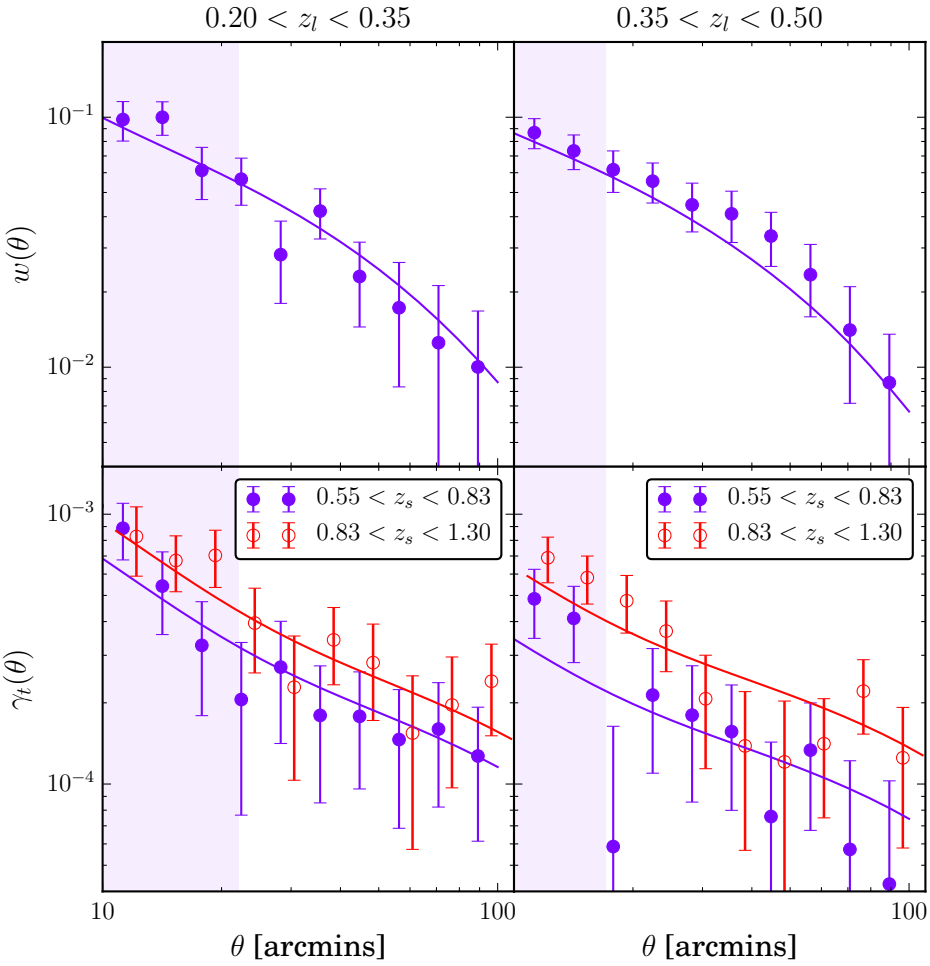

Figure 4.2: Angular galaxy clustering and galaxy-galaxy lensing measurements used in this work. For the two lens bins (left and right columns), we show the clustering measurements (upper row) and the galaxy-galaxy lensing measurements (lower row) for the two source bins, with error bars coming from jackknife resampling. The shaded region shows excluded scales in the fiducial analysis, explored in Section 4.5.1. The predictions for the best fitting curves presented in Section 4.4 are shown as the solid curves in each panel. The goodness-of-fit, as measured by the χ^2 value is 6 (3.5) for 12 (9) degrees of freedom for the high$-z$ (low-z) bin. While these reduced χ^2 values seem low, there is no evidence that the errors are overestimated to within 2-σ.

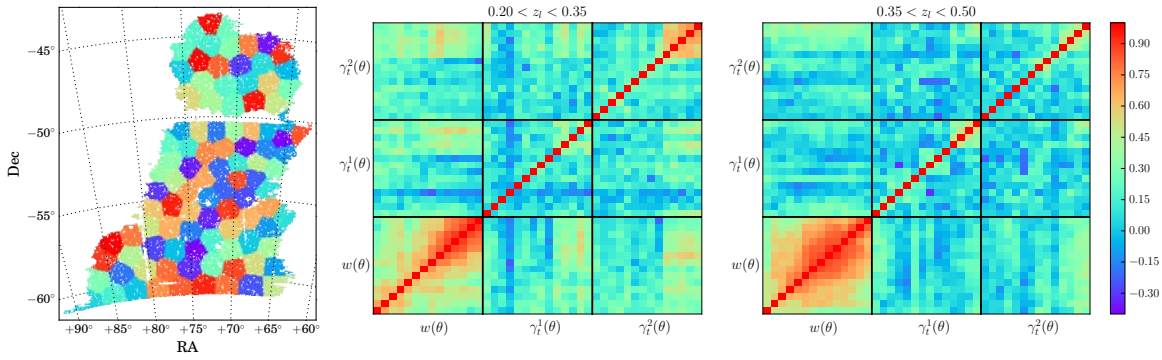

Figure 4.3: (*left panel*): DES-SV SPT-E footprint and an example of the *kmeans* jackknife regions used to compute the covariance matrices used in this work. (*center panel*): For the first lens bin, the joint jackknife correlation matrix for $w(\theta)$ and $\gamma_t(\theta)$ for the two source bins. For each submatrix of the joint correlation matrix, the angular scale ranges from 4 to 100 arcminutes in logarithmic bins. (*right panel*): Same as the *center panel*, for the second lens bin.

arcminutes. Only scales $\sim$ 20 arcminutes and above will be used in the cosmology fits (see Sec. 4.5.1 for details). The details of the calculation of the error or covariance matrix for $w(\theta)$ will be presented in Section 4.3.2.

TANGENTIAL SHEAR – $\gamma_t(\theta)$

On the lensing side, the observable is the tangential shear, i.e., the shear of the source galaxy which is perpendicular to the projected line joining the lens and source galaxies. For a given lens-source pair (j) this is given by

$$\gamma_{t,j} = -\gamma_{1,j}\cos(2\phi_j) - \gamma_{2,j}\sin(2\phi_j) \tag{4.12}$$

where $\gamma_{1,j}$ and $\gamma_{2,j}$ are the two components of shear measured with respect to a cartesian coordinate system with origin in the lens galaxy, and ϕ_j is the position angle of the source galaxy with respect to the horizontal axis of the cartesian coordinate system. Since the intrinsic ellipticity of individual source galaxies is much larger than the weak lensing shear, it is necessary to average over many such lens-source pairs. For our measurements, we compute the average in angular separation bins, θ, so that

$$\langle\gamma_t(\theta)\rangle = \frac{\sum_j \omega_j \gamma_{t,j}}{\sum_j \omega_j}, \tag{4.13}$$

where the tangential shear for each lens-source pair, j, is weighted by a factor ω_j as follows:

$$\omega_j = \frac{1}{\sigma^2_{\rm shape} + \sigma^2_{{\rm m},j}}, \tag{4.14}$$

where $\sigma_{\rm shape}$ is the shape noise intrinsic to each background galaxy, and $\sigma_{{\rm m},j}$ is the error derived from the shape measurement. We use $\sigma_{\rm shape} = 0.233$ for the `im3shape` shear catalog and $\sigma_{\rm shape} = 0.243$ for the `ngmix` shear catalog. The weights ω_j corresponding to the shear catalogs used in this work are computed and described in Jarvis et al. (2015). In order to correct for possible geometric and additive shear systematic effects, we compute the tangential shear around random lenses and subtract this from the galaxy lensing signal (as in Clampitt et al. (2016)). The result is shown in the lower panels of Fig. 4.2, over the same range of scales as for $w(\theta)$. For each lens bin we show the tangential shear using the two source bins.

4.3.2 COVARIANCES

Our measurements of $w(\theta)$ and $\gamma_t(\theta)$ are correlated across angular and source redshift bins. The joint covariance for all the measurements corresponding to each lens redshift bin is estimated from jackknife (JK) resampling, using the following expression (Norberg et al., 2009):

$$C(x_i, x_j) = \frac{(N_{\rm JK}-1)}{N_{\rm JK}} \sum_{k=1}^{N_{\rm JK}} (x_i^k - \bar{x}_i)(x_j^k - \bar{x}_j), \tag{4.15}$$

where the complete sample is split into a total of $N_{\rm JK}$ groups, x_i^k is a measure of the statistic of interest in the i-th bin using all JK regions excepting the k-th sample, and $\bar{x}_i$ is the mean of $N_{\rm JK}$ resamplings. Jackknife regions are obtained using the *kmeans* algorithm[5] run on a homogeneous random points catalog and, then, all catalogs (lenses, sources and random points) are split in $N = 100$ JK samples. *kmeans* is a clustering algorithm that subdivides n observations into N clusters (see Appendix B in Suchyta et al. 2016 for details). By applying it to a uniform random catalog with the same sky coverage as DES-SV, we define regions that are well suited for JK subsampling. The left panel in Fig. 4.3 shows our JK patches created by the *kmeans* algorithm. The resulting covariance matrices for both lens bins are also shown in Fig. 4.3 (center and right panels). The covariance is strongest between points within the $w(\theta)$ data vector. Note that: (i) we do not jointly fit both lens bins in the fiducial case so no covariances between lens bins

[5] `https://github.com/esheldon/kmeans_radec`

are shown, and, (ii) when performing cosmology fits with the lower (higher) lens bin we only use 21 (24) data points (see Sec. 4.5.1).

The JK covariance matrices shown in Fig. 4.3 contain a non-negligible level of noise. Hartlap et al. (2007) showed that the inverse of an unbiased but noisy estimator of the covariance matrix is actually *not* an unbiased estimator of the inverse covariance matrix. Therefore, when using a JK covariance matrix, a correction factor of $(N_{\rm JK} - N_{\rm bins} - 2)/(N_{\rm JK} - 1)$ should be applied to the inverse covariance, where $N_{\rm JK}$ is the number of jackknife regions and $N_{\rm bins}$ is the number of measurements (Hartlap et al., 2007). We include this correction factor in all our cosmology results.

The performance of JK covariances in DES-SV has been studied separately for galaxy clustering and galaxy-galaxy lensing in Crocce et al. (2015); Giannantonio et al. (2015) and Clampitt et al. (2016), respectively. There we generally find good agreement between true covariances from simulations or theory and the JK estimates, especially at small scales. At large scales the comparison points to an overestimation of the covariance by the JK method in the lensing case.

We have tested our method of estimating JK covariances and in particular the Hartlap correction factor, by generating a number of log-normal realizations of the convergence and matter density fields, as described in Friedrich et al. (2015). These mocks were constructed to match the lens and source galaxy densities and survey area of the DES-SV region. We used 600 mocks as an estimate of the 'true' covariance and as well as a subset of 100 mocks to represent a noisy covariance derived from independent samples. A comparable JK covariance was generated from a log-normal mock at random and dividing it into 100 patches using the same algorithm as the DES-SV data. We found that error in applying the Hartlap correction to JK samples instead of independent mocks is only a few percent compared to the total difference between using JK samples and independent samples.

In this work we also estimate the cross-covariance between galaxy clustering and galaxy-galaxy lensing, for which we find a small positive correlation among all clustering scales and large galaxy-galaxy lensing scales – the regime where the lensing errors are no longer dominated by shape noise. This is consistent with related previous work like Mandelbaum et al. (2013), where they were able to neglect this contribution due to their different noise properties. However, Marian et al. (2015) found a significant non-zero cross correlation between angular clustering and galaxy-galaxy lensing, that could contribute to biased and over optimistic constraints if ignored. As a check on the amount of covariance between probes, Fig. 4.6 also shows the result of ignoring the cross-covariance on the constraints on Ω_m and σ_8. The derived cosmology shows little deviation from our fiducial results and we find that our constraints are only minimally stronger on σ_8 (by about 3%) and weaker on Ω_m (also $\sim$3%) with a 2% improvement on $S_8 = \sigma_8(\Omega_m/0.3)^{0.16}$.

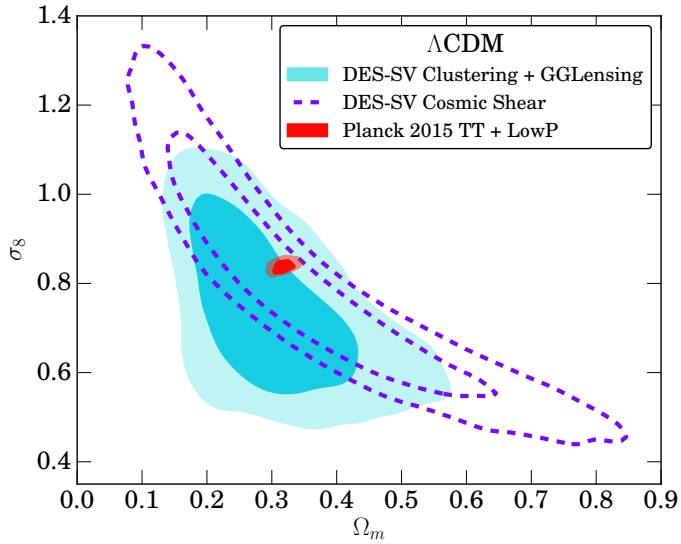

Figure 4.4: Constraints on Ω_m and σ_8 using DES-SV Cosmic Shear (dashed purple), DES-SV $w(\theta) \times \gamma_t(\theta)$ (this work, filled blue) and Planck 2015 using a combination of temperature and polarization data (TT+lowP, filled red). In each case, a flat ΛCDM model is used.

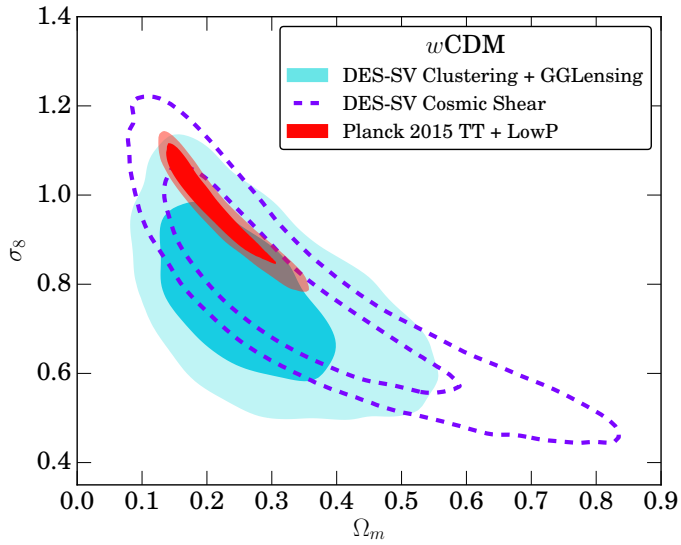

Figure 4.5: Constraints on Ω_m and σ_8 assuming a wCDM model using DES-SV Cosmic Shear (dashed purple), DES-SV $w(\theta) \times \gamma_t(\theta)$ (this work, blue) and Planck 2015 using temperature and polarization data (TT+lowP, red).

This shows that the impact of the correlation between probes is subdominant to the covariance within the same probe.

4.4 FIDUCIAL COSMOLOGICAL CONSTRAINTS

Parameter	Prior range	
Ω_m	0.1 – 0.8	Normalized matter density
Ω_b	0.04 – 0.05	Normalized baryon density
σ_8	0.4 – 1.2	Amplitude of clustering (8 h^{-1}Mpc top hat)
A_s	1.0 – 4.0 $\times 10^{-9}$	Amplitude of clustering of primordial power spectrum at pivot scale of 0.05 Mpc^{-1}
n_s	0.9 – 1.0	Power spectrum tilt
w	-5 – -0.33	Equation of state parameter
h	0.5 – 1.0	Hubble parameter ($H_0 = 100h$)
τ	0.04 – 0.12	Optical depth
b_1	1.0 – 2.2	Linear galaxy bias
b_2	-1.5 – 1.5	Next order bias parameter
β_i	-0.3 – 0.3	Shift in photo-z distribution (per source bin)
m_i	-0.2 – 0.2	Shear multiplicative bias (per source bin)
m_{IA}	-0.3 – 0.35	Intrinsic alignment amplitude (low-z source bin only)
α	-5 – -1	Additive constant $w(\theta) \rightarrow w(\theta) + 10^{\alpha}$

Table 4.1: Parameters and their corresponding priors used in this work. Not all parameters are allowed to vary in every analysis. Nuisance parameters are contained in the lower half of the table. When choosing a prior range on cosmological parameters, we allowed a sufficiently wide range to contain all of the 2-σ posterior on Ω_m, σ_8, w and h, with Planck priors on Ω_b and n_s, for which we have less sensitivity. For the systematic parameters, our choice of prior range is informed from previous DES analyses that studied the effect of shear calibration (Jarvis et al., 2015), photo-z distributions (Bonnett et al., 2015), and intrinsic alignment contamination (Clampitt et al. , 2016; The Dark Energy Survey Collaboration , 2015) on the SV catalogues. The prior on the bias parameters were taken from studies of the redMaGiC mock catalog (see Section 4.5.1 for details). In addition to the prior range on the nuisance parameters for the shear calibration and photo-z bias, there is a Gaussian prior centered around zero of width 0.5, as explained in the text.

Probes	z	σ_8	Ω_m	$S_8 \equiv \sigma_8(\Omega_m/0.3)^\alpha$	α	b_1	w_0
DES	$0.2 < z < 0.35$	0.73 ± 0.12	0.46 ± 0.12	0.77 ± 0.11	0.15	1.60 ± 0.31	-1
DES	$0.2 < z < 0.35$	0.74 ± 0.13	0.41 ± 0.14	0.77 ± 0.10	0.17	1.73 ± 0.29	-2.5 ± 1.26
DES	$0.35 < z < 0.5$	0.74 ± 0.13	0.31 ± 0.09	0.74 ± 0.12	0.16	1.64 ± 0.30	-1
DES	$0.35 < z < 0.5$	0.77 ± 0.12	0.28 ± 0.10	0.75 ± 0.11	0.13	1.71 ± 0.28	-2.03 ± 1.19
DES	$0.2 < z < 0.5$	0.76 ± 0.10	0.36 ± 0.09	0.78 ± 0.09	0.21	1.52 ± 0.28 1.60 ± 0.27	-1
Planck		0.83 ± 0.01	0.32 ± 0.01	0.82 ± 0.02	-0.49		-1
Planck		$0.98^{+0.11}_{-0.06}$	$0.21^{+0.02}_{-0.07}$	1.21 ± 0.27	-0.6		$-1.54^{+0.20}_{-0.40}$
BAO + SN + H0			0.33 ± 0.02				-1.07 ± 0.06
BAO + SN + H0 + DES	$0.35 < z < 0.5$	0.71 ± 0.1	0.32 ± 0.02	0.71 ± 0.1	0.01		-1.05 ± 0.07
DES + Planck	$0.2 < z < 0.35$	0.84 ± 0.01	0.35 ± 0.01	0.76 ± 0.02	-0.71	1.30 ± 0.13	-1
DES + Planck	$0.2 < z < 0.35$	0.89 ± 0.03	0.32 ± 0.02	0.84 ± 0.06	-0.76	1.25 ± 0.13	-1.16 ± 0.09
DES + Planck	$0.35 < z < 0.5$	0.84 ± 0.01	0.35 ± 0.01	0.76 ± 0.02	-0.71	1.41 ± 0.17	-1
DES + Planck	$0.35 < z < 0.5$	0.88 ± 0.03	0.32 ± 0.02	0.84 ± 0.06	-0.75	1.36 ± 0.14	-1.14 ± 0.09
DES + Planck + BAO + SN + H0	$0.35 < z < 0.5$	0.86 ± 0.02	0.31 ± 0.01	0.84 ± 0.03	-0.81	1.74 ± 0.28	-1.09 ± 0.05

Table 4.2: Marginalized mean cosmological parameters (and 1-σ errors) measured from the posterior distribution of a joint analysis of angular clustering and galaxy-galaxy lensing. Results for DES-SV data alone and in combination with Planck and external data (BAO, SN1a, H0) are shown for the two lens redshift bins both separately and combined. (Note that the biases are quoted separately: $b_1 = 1.52 \pm 0.28$ for $0.2 < z < 0.35$ and $b_1 = 1.60 \pm 0.27$ for $0.35 < z < 0.5$). Not shown are the additional cosmological parameters that we have marginalized, $\{n_s, \Omega_b, h_0\}$ as well as our standard set of nuisance parameters. Also quoted are the mean values and 1-σ errors given by Planck (TT+lowP) and external data alone.

In this section we present our fiducial DES-SV cosmological constraints from a joint analysis of clustering and galaxy-galaxy lensing. The data vector consists of $w(\theta)$ and the two $\gamma_t(\theta)$ measurements for the $0.35 < z < 0.5$ redMaGiC bin (see Fig. 4.2), over angular scales of 17-100 arcminutes. We chose this lens bin as our fiducial, as we estimate greater contamination from systematic errors, on both the clustering and lensing side, for the $0.2 < z < 0.35$ redMaGiC bin (see Section 4.5.6 and Clampitt et al. (2016)). To compute the model we use CAMB (Lewis et al., 2000; Howlett et al., 2012) and Halofit (Smith et al., 2003; Takahashi et al., 2012) for the linear and non-linear matter power spectra, respectively. Because the accuracy of Halofit can be confirmed only to ∼5% for certain ΛCDM models, we have checked that using the Cosmic Emulator, a more precise modelling scheme for the nonlinear dark matter power spectrum (1% to $k = 1$ Mpc^{-1}, Lawrence et al. 2010) would only affect our results at the level of ∼5% down to $10'$. We use the CosmoSIS package[6] (Zuntz et al., 2015) as our analysis pipeline and explore the joint posterior distribution of our cosmological (and nuisance) parameters using the multi-nest MCMC algorithm of Feroz et al (2009), with a tolerance parameter of 0.5, which controls the convergence of the chains, and an efficiency parameter of 0.8. Our cosmological parameters and priors are summarized in Table 4.1 and described in greater detail next in this section.

In the fiducial case, we have included two nuisance parameters per source bin (one for errors in the photo-z distribution and one for biases in the shear calibration) and one nuisance parameter per lens bin (the linear bias, b_1; the non-linear bias, b_2, accounting for scale dependence and stochasticity, is studied in Section 4.5.1), plus an additional term, α, to account for potential systematic errors induced by observational effects that might induce an overall shift in the normalisation of the amplitude of $w(\theta)$ (see Section 4.5.6). The full set of nuisance parameters and their priors are listed in the lower half of Table 4.1 and summarized below.

- **Photometric redshift calibration:** For each source bin i, we marginalize over a photo-z bias parameter, β_i, defined such as $n_i(z) \rightarrow n_i(z + \beta_i)$. In Bonnett et al. (2015), it was found that a single additive parameter for the photo-z distribution with a Gaussian prior centered on zero with a dispersion of 0.05, was sufficient to account for any statistical bias on $\Sigma_{\rm crit}$ and hence σ_8 within the degree of statistical error expected for the SV catalogs.

- **Shear calibration:** For each source bin i, we marginalize over an extra nuisance parameter m_i, to account for the shear calibration uncertainties, such that $\gamma_{t;i}(\theta) \rightarrow (1 + m_i)\gamma_{t;i}(\theta)$, with a Gaussian prior with mean 0 and

[6] `https://bitbucket.org/joezuntz/cosmosis`

width 0.05, as advocated in Jarvis et al. (2015).

- **Additive $w(\theta)$ constant:** We marginalize over an additive constant parameter, α, in the galaxy angular correlation function: $w(\theta) \rightarrow w(\theta)+10^{\alpha}$. This parameter accounts for possible systematics arising from variations in observing conditions across the field, stellar contamination and masking (Ross et al., 2011), which we also test for in the next section.

Probes	z	$100\beta_1$	$100\beta_2$	$100m_1$	$100m_2$	α
DES (ΛCDM)	$0.2 < z < 0.35$	-0.89 ± 4.58	0.25 ± 4.56	-0.09 ± 4.59	0.44 ± 4.42	-3.41 ± 0.84
DES (wCDM)	$0.2 < z < 0.35$	-1.00 ± 4.53	0.13 ± 4.51	-0.85 ± 4.47	0.14 ± 4.57	-3.42 ± 0.83
DES (ΛCDM)	$0.35 < z < 0.5$	-1.77 ± 4.46	0.14 ± 4.67	-0.05 ± 4.65	0.36 ± 4.64	-3.57 ± 0.81
DES (wCDM)	$0.35 < z < 0.5$	-1.78 ± 4.38	0.18 ± 4.48	-0.85 ± 4.48	0.05 ± 4.31	-3.49 ± 0.81
DES + Planck (ΛCDM)	$0.2 < z < 0.35$	-0.58 ± 4.83	0.29 ± 4.99	-0.63 ± 4.87	0.72 ± 4.84	-3.62 ± 0.82
DES + Planck (wCDM)	$0.2 < z < 0.35$	-0.87 ± 4.73	0.14 ± 4.87	-0.76 ± 4.88	0.41 ± 4.79	-3.62 ± 0.82
DES + Planck (ΛCDM)	$0.35 < z < 0.5$	-3.11 ± 4.48	-0.53 ± 4.95	-0.99 ± 4.92	-0.65 ± 4.77	-3.44 ± 0.87
DES + Planck (wCDM)	$0.35 < z < 0.5$	-1.04 ± 2.53	-0.16 ± 2.64	-1.09 ± 4.32	-0.68 ± 4.34	-3.43 ± 0.85

Table 4.3: Marginalized mean systematic uncertainty parameters with 1-σ errors measured from the posterior distribution of the joint analysis of angular clustering and galaxy-galaxy lensing in DES-SV data. We assume a Gaussian prior (centered on zero) for each systematic parameter, while the width of the prior is set from Jarvis et al. (2015) for the shear calibration and Bonnett et al. (2015) for the photo$-z$s. Each nuisance parameter is additionally truncated by the amounts in Table 4.1.

The resulting constraints in the Ω_m and σ_8 plane are shown in Fig. 4.4. The 2D contours are centered around $\Omega_m \sim 0.3$ and $\sigma_8 \sim 0.75$, and marginalizing out the other parameter we find the following 1D constraints: $\Omega_m = 0.31 \pm 0.10$ and $\sigma_8 = 0.74 \pm 0.13$. Comparing to measurements from Planck (The Planck Collaboration et al. , 2015) and DES Cosmic Shear (The Dark Energy Survey Collaboration , 2015) alone, we are consistent at the $\sim 1\sigma$ or better level. We combine results from the two experiments in Section 4.6. In addition, we see the same direction of degeneracy between these two parameters as with cosmic shear, although the degeneracy is not quite as strong with $w(\theta)$ and $\gamma_t(\theta)$.

We also include w, the dark energy equation of state parameter, as an additional free parameter in Fig. 4.5. We found that the DES-SV data alone was unable to provide strong constraints on w and obtained $w = -1.93 \pm 1.16$. However, compared to Planck (red contours), the DES-SV constraints on Ω_m and σ_8 are degraded far less when w is introduced as a free parameter. Also, we note that the preference for $w < -1$ values is determined by our choice of prior on w; we require $-5 < w < -0.33$, so the prior volume covered by $w < -1$ is greater than $w > -1$ and in the absence of a strong constraint on w, values of $w < -1$ are favored.

Table 4.2 contains a more detailed summary of our findings for this fiducial setup, assuming either a ΛCDM or wCDM cosmology. In addition to DES $w(\theta)$ and $\gamma_t(\theta)$, we show results combined with Planck. Table 4.2 also shows results for our lower redshift lens bin, $0.2 < z < 0.35$. For these results we vary only the cosmological parameters $\{\Omega_m, \Omega_b, h, n_s, \sigma_8\}$ and w where noted (in addition to the nuisance parameters described in the present and following sections). When combined with constraints from Planck, we also allow the optical depth, τ, to vary as well, since the CMB has additional sensitivity to physics that is only weakly captured by large scale clustering at late times and we fit for the amplitude of the primordial power spectrum, A_s, rather than σ_8. Table 4.3 shows the constraints on the nuisance parameters related to photo-z and shear calibration described above.

In the following section, we will study the robustness of these results under changes in the configuration of the data vector and the systematics modelling.

4.5 ROBUSTNESS OF THE RESULTS

In this section, we describe the suite of tests performed to check that our conclusions are unbiased with respect to errors in the shear and photo-z calibrations, intrinsic alignments, survey geometry, choice of angular scales and theoretical modelling of the data vectors. The results in this section are displayed in Fig. 4.6,

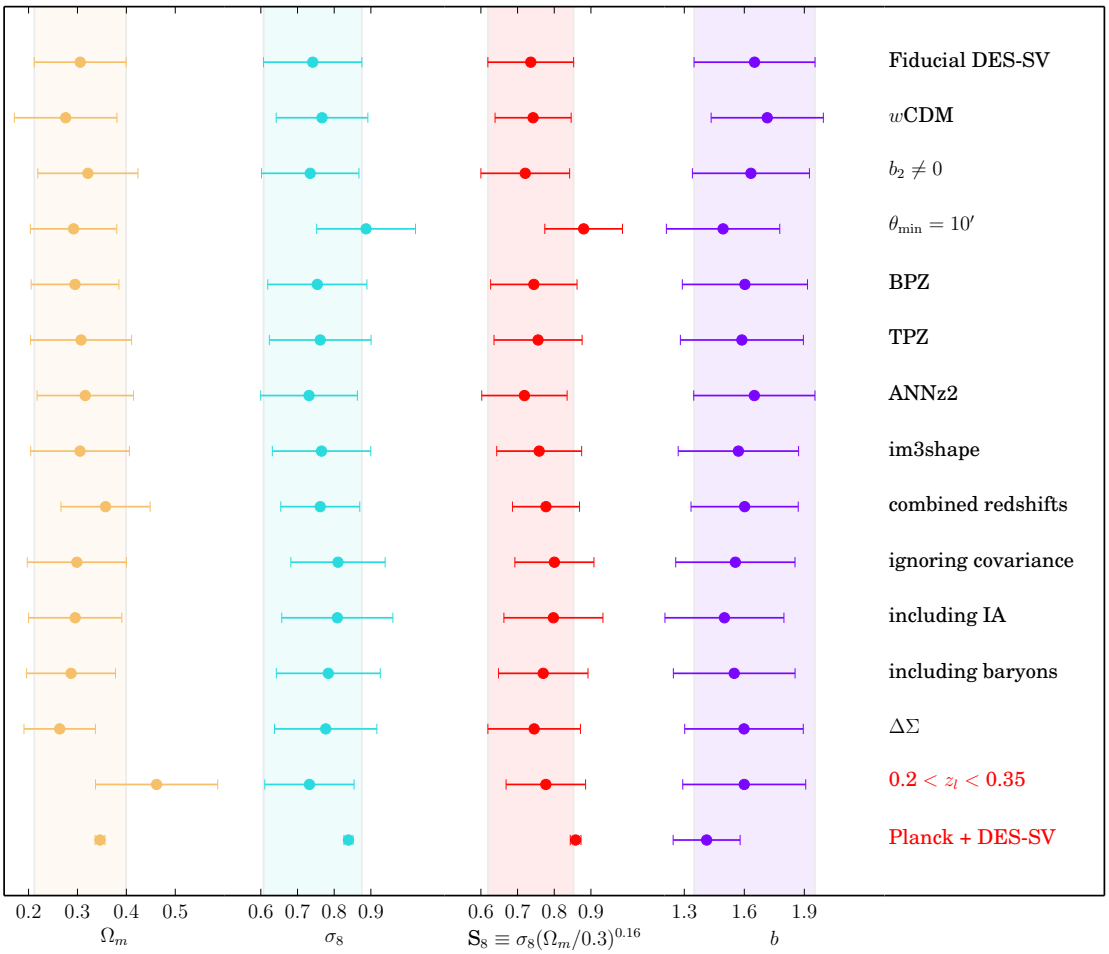

Figure 4.6: Marginalized 1D posterior constraints on $\{\Omega_m, \sigma_8, S8, b_1\}$ for the lens bin $0.35 < z < 0.5$ for various configurations in our pipeline. For this figure, we have defined $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.16}$, that is, we hold the index fixed to the degeneracy direction found for our fiducial analysis. Note that this value is substantially different to one favoured by Planck data alone, but we have chosen a constant value to enable comparisons between the systematic tests. Our fiducial results use shear catalogs from `ngmix`, SkyNet photometric redshifts, and linear bias in a ΛCDM cosmology, as described in Section 4.4. The different rows in this plot are obtained by varying the fiducial assumptions individually to test their impact on the parameter constraints, and they are all detailed in Section 4.5 and the Appendix. Tests involving (nearly) independent data are highlighted in red near the end of the table.

for the parameters we are most sensitive to in this work: $\{\Omega_m, \sigma_8, b_1\}$. The different rows correspond to the different tests described in this section or in the Appendix, where we check the results from a different lensing estimator. Despite the changes in the photo-z algorithms, the shear catalogs, the weighting of the lens-source pairs, non-linear bias modelling and choice of scale, our estimates for these cosmological parameters in Fig. 4.6 usually remain within 1-σ of the fiducial constraints.

A number of systematics that are unique to the measurement of the tangential shear such as the calibration of galaxy ellipticities, the effect of different shear calibration pipelines, null detection of the cross component and effect of photo-z errors in the lens and source catalogs on the measurement have already been accounted for in Clampitt et al. (2016), so we do not present tests for these effects again. For more information on tests of the shear pipeline, we refer the reader to Jarvis et al. (2015) while Bonnett et al. (2015) contains extensive tests of the photo-z calibration algorithms. We also check for possible systematics introduced by the effects of survey geometry, depth and varying observing conditions in the survey following the techniques in Crocce et al. (2015).

Our analysis pipeline accounts for the effect of a number of systematics which are folded into our final constraints on cosmology. To first order, these nuisance parameters are responsible for altering the amplitude of $w(\theta)$ and $\gamma_t(\theta)$, and so are strongly degenerate with one another. As a result, we were unable to constrain these parameters beyond their prior distributions and the results in Table 4.3 show that the posterior distributions of the nuisance parameters no more informative than the priors. To determine which of these most affect our results, we have analysed each of these systematics individually by running chains in four scenarios: no systematics, shear calibration only, photo-z errors only, full weak lensing systematics but no constant offset in w(θ), and shear calibration with photo-z errors (our fiducial set up). We found that including an additive constant to $w(\theta)$ was responsible for the greatest decrease in precision on the 1D marginalized constraints on Ω_m, with the 1-σ error on Ω_m increasing by as much as 17% compared to the no systematics case. However, σ_8 was much less affected with a difference below 3%. In comparison, accounting for photo-z errors with an additional two free parameters in the $N(z)$ distribution increased the error on both parameters by about 8%. The change from including two shear calibration parameters was smaller still, with only a 3% reduction in precision for Ω_m and 5% for σ_8 relative to the no systematics case. We also found small changes to the best fitting values, well within the 1-σ confidence interval, as expected from Fig. 4.6.

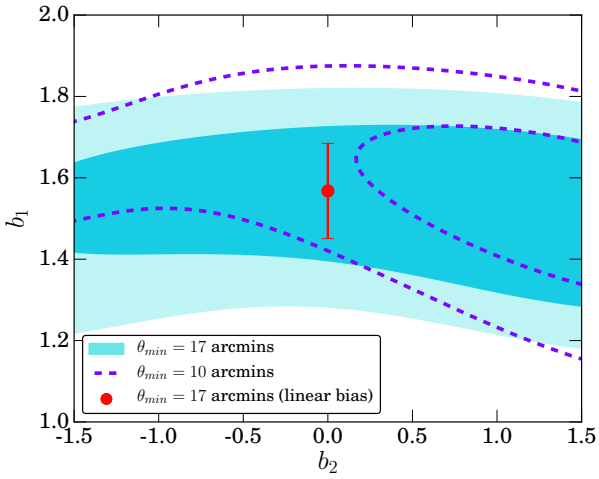

Figure 4.7: The posterior distribution on the bias parameters, b_1, b_2 from simulations of $w(\theta)$ for the redshift bin $0.35 < z < 0.5$. We fit the McDonald (2006) model to a minimum cut in scale at $10'$ (cyan) and $17'$ (purple) and a linear bias model to $17'$ (red point) to demonstrate the insensitivity of our fiducial results with a $17'$ cutoff to b_2.

4.5.1 CHOICE OF SCALES

There are several reasons to limit the range of scales that we consider in our analysis. The large scale cutoff is set by the size of the SV patch and how well the geometry of the region can be modelled; we found that our jackknife estimates of the covariance matrix overestimated the covariance matrix obtained from 50 independent N-body simulations above $70'$ (see Fig 5, Clampitt et al. 2016).

On small scales, we are limited by how well we can model the nonlinear clustering of matter and of redMaGiC galaxies. Galaxy formation preferentially occurs in high density environments within dark matter halos and is subject to a number of complex baryonic processes; these are not captured in our model predictions for the mass power spectrum and potentially introduce a non-trivial bias between the dark matter and the galaxies. This is particularly important for the tangential shear, which contains a mixture of small and large scale information; i.e. imposing a sharp cutoff in angular scale does not completely eliminate the effect of scales below that cutoff (Mandelbaum et al., 2013). On small enough scales, we expect to observe effects such as stochasticity, non-local bias and scale dependence. These could invalidate the linear bias model used in our analysis.

In this section, we present simulation based tests to determine the smallest scales for which the linear bias model and perturbation theory model of McDonald (2006) are valid. We use a mock catalog designed to reproduce the properties of the DES-SV survey. The catalog is based on an N-body simulation (c-400; see also Mao et al. (2015); Lehmann et al. (2015)) run with the L-GADGET code, a variant of GADGET (Springel , 2005). The simulation has a box size of 400 Mpc/h with 2048^3 particles and a force resolution of 5.5 kpc/h. Halo catalogs were generated with the ROCKSTAR halo finder (Behroozi et al. , 2013a) and the CONSISTENT TREES merger tree builder (Behroozi et al. , 2013b). A galaxy catalog

was produced using an abundance matching technique, as described in Reddick et al. (2013) and Lehmann et al. (2015), with halos ranked according to the peak halo velocity and assigned a luminosity from the Blanton et al. (2003) luminosity function, using a scatter of 0.2 dex. Snapshots from the simulation were combined into a lightcone with the same footprint as the DES-SV region. Galaxy colors were assigned using the empirically derived relationship between luminosity, projected distance to the fifth nearest neighbor galaxy, and galaxy SED (this method for assigning colors has been used in previous generations of catalogs, see e.g. Cunha et al. 2012; Chang et al. 2015). Photometric errors were added to match the depth distribution of DES-SV galaxies. The redMaGiC algorithm was run on the lightcone, using the same technique as applied to the DES-SV data and this produced a mock redMaGiC catalog. The redMaGiC color model is retuned to the simulations before identifying these galaxies, but was found to have similar properties to that seen in the data. We find that the clustering properties of the redMaGiC galaxies in this catalog are consistent with those measured in DES-SV data.

From the mock catalog, we have measured $w(\theta)$ in the same bins in redshift, $0.2 < z < 0.35$ and $0.35 < z < 0.5$, from $10' < \theta < 100'$. Our covariance matrix is calculated from a jackknife resampling of the catalog as described in Section 4.3.2.

We test our bias modelling by making two cuts in angular scale at $10'$ and $17'$, corresponding to ($\sim$ 3 Mpc/h) and ($\sim$ 5.5 Mpc/h), because we expect the bias to transition between its large scale asymptotic limit to scale dependence somewhere in this regime for the galaxy type that we consider. We fit both a linear and a quasilinear bias model with two free parameters, b_1 and b_2, as described in Section 4.2.1 to the simulated $w(\theta)$ while holding the cosmological parameters fixed to the value of the N-body simulation. Note that the effect of the shot noise parameter, N, on $w(\theta)$ is negligible on our scales of interest so we do not include it in our tests. Figure 4.7 shows the recovered biases when all the cosmological parameters are fixed at the simulation values for the fiducial lens bin of $0.35 < z < 0.5$. The measured $w(\theta)$ is insensitive to the value of b_2 when a minimum angular scale of $17'$ is chosen (cyan filled contour) and we are simply recovering our prior distribution on b_2.

When we change the minimum scale to $10'$ (purple dashed contour), there is a 1-σ preference for a non-zero value. Using a linear model of biasing (Fig. 4.7; red point) with the same fixed cosmology set up, we find that we recover the same value of b_1 as in the non-linear case. We obtain similar results for the low-z lens bin, except that the minimum scale cutoff is now at $22'$ for $w(\theta)$ to be well modelled by a linear bias. Figure 4.7 demonstrates that our choice of using a linear bias up to these angular scales for the redMaGiC sample should not affect our ability to constrain cosmology. Based on these results, we can conclude that

applying a linear bias model with $\theta_{\rm min} = 17'$ ($22'$) for the high$-z$ (low-z) lens bin will not bias our results in the presence of scale dependent non-linear biasing. Since our simulations do not address the lensing component of our analysis, we have performed to additional checks using the data. We have rerun our fiducial analysis with b_2 as an additional free parameter, while keeping $\theta_{\rm min} = 17'$. For these fits, we obtained $\Omega_m = 0.32 \pm 0.10$, $\sigma_8 = 0.73 \pm 0.13$, $b_1 = 1.63 \pm 0.29$ and $b_2 = -0.14 \pm 0.76$, which is consistent with our fiducial results. We have also tested our small scale cut off by using $\theta_{\rm min} = 10'$ with our fiducial set up. We found that this increased the value of σ_8 to 0.887 ± 0.134 from $\sigma_8 = 0.741 \pm 0.134$. Since the decrease in the error bars on our cosmological parameters of interest is not significant, we kept our existing value of $\theta_{\rm min}$ to 17'.

For the shear catalogues, Jarvis et al. (2015) identified $3'$ as the angular scale in the shear auto correlation function at which the additive errors contribute to half of the total forecasted error on the measurement of σ_8 or about $\sim 3\%$. Although it is expected that position-shear correlations are less sensitive to additive systematics in the shear, we only consider angular scales $\theta \geq 10'$ even for the tests of the bias model above. This $3'$ cutoff is well outside of the minimum scales considered in our cosmological analysis which use at most $\theta > 17'$.

4.5.2 PHOTO-z SYSTEMATICS

Since DES-SV is an imaging survey, the quality of our constraints rely heavily on being able to robustly calibrate the photometric redshifts of the lens and source galaxy samples. However, because $w(\theta)$ does not use radial information, apart from the selection function, it is relatively insulated from photometric errors compared to the full 3D correlation function. Furthermore, because the photometric error in the lens redMaGiC sample is so small (Rozo et al., 2015b), the potential systematic errors in the cosmology analysis are dominated by the photometric redshifts of the source galaxy sample.

We deal with photometric redshift systematics in two different ways. First, we follow the recommendations of Bonnett et al. (2015) and define an additional photo-z bias parameter for each source bin, i, as:

$$n_i^{pred}(z) = n_i^{obs}(z + \beta_i) \tag{4.16}$$

where β_i is a free parameter with a Gaussian prior of width 0.05 to be constrained during the fitting process. The width of the prior is set to be consistent with Bonnett et al. (2015), where it was found that the difference between photometric and spectroscopic estimates of the redshift of the training samples that most closely resemble our shear catalogs have a relative mean bias with a Gaussian dispersion

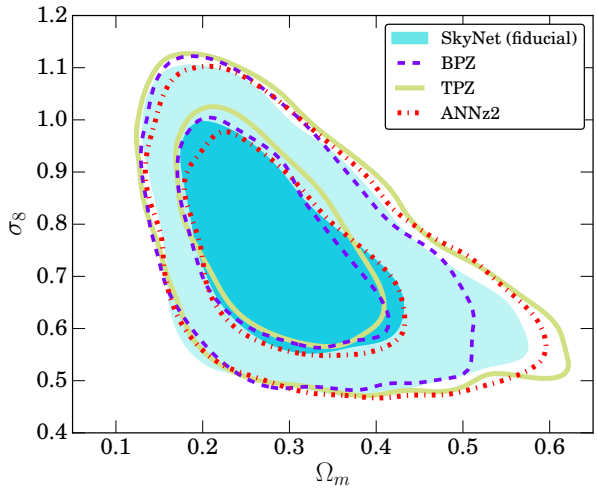

Figure 4.8: Constraints on Ω_m and σ_8 using four different photo-z codes to calculate the redshift distribution of sources. The contours for the $0.35 < z < 0.5$ redshift bin are shown here; we have also checked the robustness of our results for lenses in the redshift bin $0.2 < z < 0.35$.

of 0.05. This method was also used in the DES-SV Cosmic Shear Cosmology paper (The Dark Energy Survey Collaboration , 2015). We found that introducing an additional photo-z bias parameter for each source bin increases our uncertainty by, at most, 8% compared to the constraints we would have if we did not fit for any systematic parameters.

In addition, we check that our constraints are robust to our choice of photo-z algorithm. Our fiducial shear catalogs use photometric redshifts derived from the SkyNet algorithm (Graff et al., 2014; Bonnett et al., 2015), and we have repeated our analysis by using the redshift distribution given by three other photo-z codes studied in Bonnett et al. (2015), namely BPZ, TPZ and ANNz2. For this test, we assume a ΛCDM cosmology and allow the cosmological parameters $\{\Omega_m, \Omega_b, h, \sigma_8, n_s, b_1\}$ to vary. In addition, we also fit for the usual systematic parameters, β_i for the photo-z bias and m_i for the multiplicative bias in the shear calibration and the same prior distributions. The resulting constraints in Fig. 4.8 (and 4.6) show that our results are insensitive to the choice of the photo-z algorithm.

Interested readers should refer to Bonnett et al. (2015) for a full discussion of the photo-z methods considered and the systematics modelling that we have only summarized here.

4.5.3 SHEAR CALIBRATION SYSTEMATICS

Here we present our approach to modelling a possible residual error in the shear calibration. For the interested reader, the full details of the production and testing of the shear catalogs used in this analysis are given in Jarvis et al. (2015).

Similar to the photometric redshift case, we deal with potential shear calibration systematics on two fronts. Firstly, we include an extra nuisance parameter

for the shear calibration, m_i, as:

$$\gamma_{t;i}^{pred}(\theta) = (1 + m_i)\,\gamma_{t;i}^{obs}(\theta) \tag{4.17}$$

with a Gaussian prior, $p(m_i)$, with mean 0 and width 0.05, for each source bin i in our analysis as recommended in Jarvis et al. (2015). Contamination from additive errors in the shear estimation are expected to be minimal for galaxy-galaxy lensing, because of the azimuthal symmetry of the lens system. Including an additional parameter for the shear calibration degrades our constraints by, at most, 5%, compared to all systematic parameters being ignored or set to fixed values.

Secondly, galaxy images in the DES-SV region were analyzed with two pipelines, `ngmix` and `im3shape`. Jarvis et al. (2015) showed that they both produced consistent results that satisfied the SV requirements for weak lensing, i.e. that less than half of the forecasted error on σ_8 (about 3%) originates from systematics in the measurement of the shear. Although we have chosen to use the `ngmix` catalog for our analysis, we have also rerun the analysis pipeline on the `im3shape` catalog to check that our results are not sensitive to the shear catalog used (see Fig. C.1 for a comparison of lensing measurements using the two shear pipelines). We found that the cosmological parameters varied imperceptibly when the `im3shape` catalog was used instead of `ngmix`. This is demonstrated in Fig. 4.6.

4.5.4 INTRINSIC ALIGNMENTS

Correlations between the intrinsic shapes and orientations of lensing sources, known as "intrinsic alignments" (IA), are one of the most significant astrophysical sources of uncertainty in weak lensing measurements (see Troxel & Ishak 2015; Joachimi et al. 2015 for recent reviews). Although typically considered in the context of shear-shear correlations, IA can also contaminate galaxy-galaxy lensing measurements due to uncertainties in photo-z estimates which lead to overlap in the true lens and source distributions (see Fig. 4.1). The intrinsic shapes of sources can be correlated with the positions of lenses at the same redshift (Blazek et al. , 2012).

In general, the contamination from IA reflects the (potentially nonlinear) relationship between large-scale structure and galaxy shapes, as well as the clustering of lenses and physically associated sources. However, observational evidence (e.g. Joachimi et al. 2011; Blazek et al. 2011; Singh & Mandelbaum 2015) indicates that the dominant IA contribution is likely from elliptical (pressure-supported) galaxies, for which the IA component is linearly related to the large-scale tidal field. This "tidal alignment" paradigm (Catelan et al. , 2001; Hirata & Seljak ,

2004; Blazek et al. , 2015) was recently used to mitigate IA in the DES-SV Cosmic Shear Cosmology analysis (The Dark Energy Survey Collaboration , 2015). In this work, we consider scales on which the clustering of lens-source pairs is negligible (see Clampitt et al. 2016 for further discussion). In this regime, tidal alignment predicts that the fractional IA contamination to the lensing signal is nearly scale-invariant. Both the IA and lensing are sourced by the same matter power spectrum, even in the presence of nonlinear evolution, and we find that the different line-of-sight weighting for IA and lensing (e.g. Eq. 4.3) leads to negligible relative scale-dependence in angular correlations.

We thus account for the potential impact of IA in our analysis by including an additional term that modifies the amplitude of the tangential shear, such that $\gamma_t(\theta) \rightarrow (1 + m_{\text{shear cal}} + m_{\text{IA}})\gamma_t(\theta)$. We place a Gaussian prior on m_{IA} of 8% $\pm$4% for the lower redshift source bin, corresponding to the IA amplitude constraint of approximately $A_{IA} = 2 \pm 1$ from the cosmic shear analysis of the same sources on the DES-SV patch (The Dark Energy Survey Collaboration , 2015). The same calculation indicates that the higher redshift source bin is sufficiently separated from the redshift of the lenses that the potential IA contamination is negligible. Potential IA contamination in the galaxy-galaxy lensing measurement is discussed further in Clampitt et al. (2016).

We do not observe a significant detection of IA contamination beyond the prior imposed; we find that $m_{\text{shear cal,1}} + m_{\text{IA,1}} \sim 8.0 \pm 3.7\%$ for the low redshift sources with $m_{\text{shear cal,2}} \sim -5.3 \times 10^{-4} \pm 4.5\%$ for the higher source bin. Including IA only affects the cosmology results by, at most, inducing a $\sim$3% shift towards a lower value of Ω_m compared to the fiducial case without IA, as shown in Fig 4.6. For σ_8, the change was much smaller, with a fractional shift of less than a percent. Because the inclusion of IA contamination has a negligible effect on our results, compared to the statistical errors, we do not include IA modelling for our fiducial analysis.

4.5.5 IMPACT OF BARYONS

One of the most challenging sources of systematic error affecting weak lensing results is the impact of baryonic effects on small scale clustering. Much of the behavior of baryonic content on small scales is unknown; the nonlinear clustering can only be modelled by computationally expensive N-body simulations but the results vary greatly with the simulation parameters such as the amount of supernova and active galactic nuclei (AGN) feedback allowed. To minimize the impact of baryonic effects on our results, we choose to truncate our measurements conservatively to large scales. However, we have implemented a scheme for evaluating the effect of baryons on our results similar to that used for the DES cosmic

shear analysis The Dark Energy Survey Collaboration (2015). We take the power spectrum measured from the OWLS simulation (van Daalen et al. , 2011) with AGN feedback, since this model induces the most extreme changes to small scale clustering while also matching results of X-ray and optical observations (McCarthy et al. , 2011), and replaced the dark matter power spectrum evaluated by Halofit thusly:

$$P(k,z) \rightarrow \frac{P_{\rm AGN}}{P_{\rm DM}} P(k,z) \tag{4.18}$$

where $P_{\rm AGN}$ and $P_{\rm DM}$ are the power spectra measured from AGN and dark matter only models from the OWLs simulations respectively. We found that including baryonic effects in this way, affects our results by a negligible amount as shown in Fig. 4.6 and the constraints shift to $\sigma_8 = 0.784 \pm 0.14$ and $\Omega_m = 0.287 \pm 0.09$ from $\sigma_8 = 0.741 \pm 0.13$ and $\Omega_m = 0.306 \pm 0.09$. For our fiducial results, we have chosen to ignore the impact of baryons.

4.5.6 IMPACT OF OBSERVING CONDITIONS

Photometric galaxy surveys such as DES are affected by time-dependent fluctuations in observing conditions that may impact the galaxy catalogs. There are a number of effects that can modulate the detection efficiency of galaxies and cause density variations across the survey footprint. In this section we follow the approach of Crocce et al. (2015) and consider single-epoch properties that affect the sensitivity of the survey and hence may affect the galaxy clustering and galaxy-galaxy lensing observables. We use projected HEALPix[7] (Górski et al., 2005) sky maps (with resolution `nside`=4096) in $griz Y$ bands for the following quantities:

- `depth`: mean survey depth, computed as the mean magnitude for which galaxies are detected at $S/N = 10$.

- `FWHM`: mean seeing, in pixel units, computed as the full width at half maximum of the flux profile.

- `airmass`: mean airmass, computed as the optical path length for light from a celestial object through Earth's atmosphere (in the secant approximation), relative to that at the zenith for the altitude of CTIO.

- `skysigma`: mean sky background noise, computed as the flux variance per amplifier in chip of the CCD.

[7] `http://healpix.sf.net`

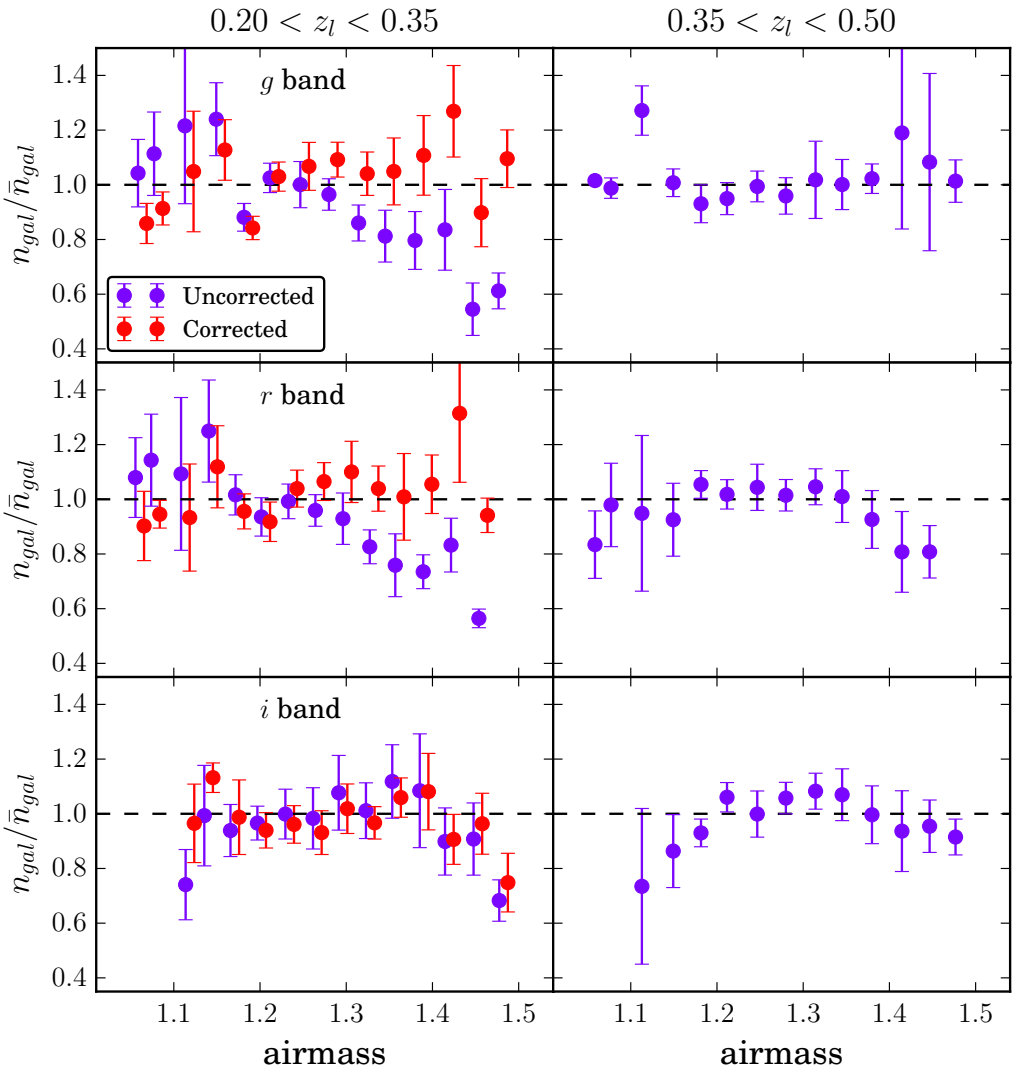

Figure 4.9: redMaGiC galaxy density as a function of `airmass` in g, r and i bands for the two lens redshift bins considered in this work. A significant correlation is present for the g and r bands in the low-z bin, which we correct by weighting galaxies inversely by the `airmass` values at the sky position. Note that we do not apply a correction to the high-z bin since it doesn't show a significant correlation with any systematics parameter.

- `USNO`: mean stellar density, as measured by the USNO-B1 stellar catalog (Monet et al., 2003) with B magnitude brighter than 20 to ensure constant depth across the field.

See Leistedt et al. (2015) for a full description of these maps.

We study the density of redMaGiC galaxies in the two lens bins as a function of each of these quantities that can potentially result in systematic effects. To ensure the data is free of such systematics, we require the galaxy density to be uncorrelated with the observed `depth`, `FWHM`, `airmass`, `skysigma` and `USNO`, otherwise we apply a correction to remove the dependency. Among the five quantities for each band and each lens bin considered here, we only find a significant correlation in the low-z bin with `airmass` in the g and r DES bands. This trend is demonstrated in Fig. 4.9, which shows the redMaGiC galaxy density as a function of `airmass` in g, r and i bands for the two lens bins. In order to correct for this correlation, we weight galaxies according to the inverse of a linear fit to the observed trend of `airmass` in the g band. This procedure is similar to that applied in Ross et al. (2012, 2014) to correct for systematic relationships with stellar density and `airmass`. The corrected results are shown in Fig. 4.9, where we see that the g band weighting also corrects the trend in the r band, as expected given the correlation present among the `airmass` maps in the g and r bands.

In addition to the weighting correction described above, we have also applied the procedure used in Crocce et al. (2015), in which galaxy and systematics maps are cross-correlated and used to correct the galaxy correlation functions. At the galaxy clustering level, the two approaches yield consistent results. Furthermore, in both cases the correction is compatible with an additive constant in the angular galaxy clustering signal. Nonetheless, we introduce an additive constant as a systematics parameter in the corrected measurement of $w(\theta)$ as outlined in Section 4.4 to deal with any residual systematic effects. This is marginalized over in the cosmological analysis according to the prior defined in Table 4.1. On the other hand, the impact of the `airmass` correction in the galaxy-galaxy lensing observables is not significant given the statistical power of these observations in DES-SV.

As opposed to Crocce et al. (2015) we do not find the `depth` and `FWHM` maps to be relevant for our lens sample, mainly because redMaGiC galaxies are much brighter than the DES main galaxy sample (Benchmark) considered in that work. On the other hand, correlations between `airmass` maps and galaxy positions were not found to be a significant systematic in Crocce et al. (2015), while for redMaGiC galaxies in the low-z lens bin, this was the only observing condition with a substantial impact on clustering. While Crocce et al. (2015) includes all types of galaxies, the redMaGiC selection process preferentially chooses red galax-

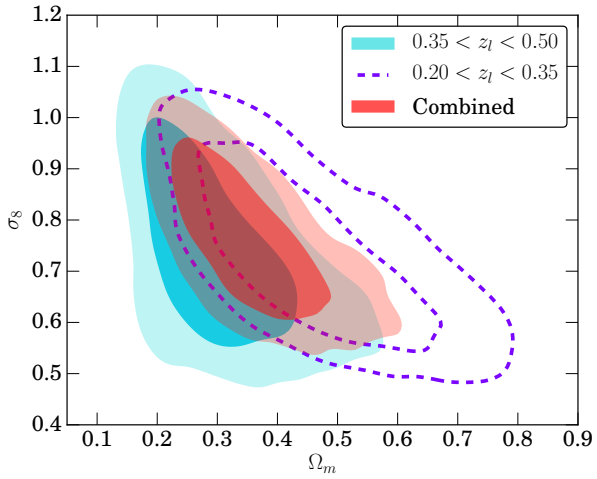

Figure 4.10: Constraints on Ω_m and σ_8 using DES-SV $w(\theta) \times \gamma_t(\theta)$. The fiducial high-$z$ lens bin is shown in filled blue, the low-z lens bin is shown as dashed purple lines and the combination of the two lens bins is shown in filled red. In each case, a flat ΛCDM model is assumed.

ies as described in Section 4.3. It is plausible that these galaxies are more affected by `airmass`, via their sensitivity to atmospheric extinction. At high `airmass`, the filter bandpasses shift to the red and the RedMapper color selection, in which redMaGiC relies, do not compensate for this. The effect is more important for the bluer DES bands g and r (Li et al., 2016), and the key spectral features of red galaxies, like the 4000Å break, fall in a bluer window of the filter set at lower redshifts, and hence the effect of atmospheric extinction is enhanced for our low-z lens bin.

In the following subsection, we present cosmology results with the low-z lens bin after correcting for the correlation with `airmass`.

4.5.7 LOW-z LENS BIN RESULTS

In this section we present the cosmology results obtained for the low-z redMaGiC lens bin ($0.20 < z < 0.35$), described in Section 4.3.1 and for which measurements are shown in Fig. 4.2. For this bin, a significant correlation of the galaxy density with `airmass` was found and corrected for in Section 4.5.6.

The photo-z and shear systematics treatment in the cosmology pipeline is equivalent to that of the fiducial lens bin and we use these results as another robustness check for the cosmological analysis performed in this work.

The cosmological constraints obtained from these measurements are shown in Fig. 4.6 and Table 4.2, and the constraints on Ω_m and σ_8 from the combination with the fiducial high-z lens bin are shown in Fig. 4.10. For most of the parameters, these lower redshift lenses are in agreement with our fiducial setup, but Ω_m shows a preference for higher values after correcting for the observing conditions described in Section 4.5.6. Still, the results for both lens bins are within 1σ of each other.

Having confirmed that the results from both the low and high redshift lens bins are consistent, we explore fitting them jointly in the same analysis pipeline to improve our constraints on cosmology. The covariance between lens bins may include a contribution from shape noise in the shear catalog. We estimate this contribution by introducing a random direction to the measured ellipticities before calculating the tangential shear. This is performed ∼300 times to obtain a jackknife estimate of the shape noise across lens and source bins. We then add the shape noise as an off-diagonal component to the covariance matrix between lens bins with the diagonal components being the usual JK covariance matrices used for individual fits. We find that the marginalized constraints are $\Omega_m = 0.36 \pm 0.09$ and $\sigma_8 = 0.76 \pm 0.11$, which show very little improvement on our fiducial results. However, the constraint on $S_8 \equiv \sigma_8(\Omega_m/0.3)^\alpha$, where α is chosen to be perpendicular to the degeneracy direction in the Ω_m-σ_8 plane, shows a reduction in the error, from $S_8 = 0.735 \pm 0.117$ ($\alpha = 0.16$; high-z lenses only) to $S_8 = 0.782 \pm 0.088$ ($\alpha = 0.21$; all lenses). These values of Ω_m, σ_8 and S_8 are shown in Fig. 4.6. We do not however consider this arrangement as our 'fiducial' model, leaving joint constraints to future work with additional survey area.

4.6 DISCUSSION

We have presented our baseline cosmological results from DES data in Section 4.4, assuming a flat ΛCDM model in Figure 4.4 and a flat wCDM model in Figure 4.5. Our results for the marginalized mean parameter values are contained in Table 4.2 for each lens bin, with and without external data sets. We also show results for each of the nuisance parameters used in our fits in Table 4.3.

4.6.1 EXTERNAL DATASETS

We performed a joint analysis of our measurements with the Planck 2015 temperature and polarization auto and cross multipole power spectra, $C^{TT}(\ell)$, $C^{TE}(\ell)$, $C^{EE}(\ell)$ and $C^{BB}(\ell)$. Specifically, we use the full range of $C^{TT}(\ell)$ from $29 < \ell < 2509$ and the low-ℓ polarization data from $2 < \ell < 29$, which we denote as Planck (TT-lowP). The inclusion of the maps allows for stronger constraints on τ which in turn affects A_s, the primordial power spectrum amplitude. We have also chosen this configuration to allow for an easy comparison with the DES-SV Cosmic Shear Cosmology paper (The Dark Energy Survey Collaboration , 2015). The constraints from only using this configuration of Planck data when assuming a wCDM model are shown as the red contours in Fig. 4.5.

With the inclusion of the DES $\gamma_t(\theta)$ and $w(\theta)$ measurements, we were able to improve on the constraints on σ_8 and w from just Planck alone, which prefers

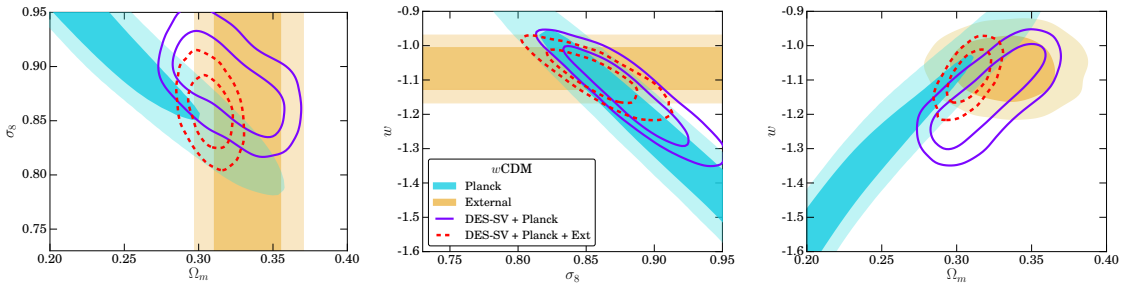

Figure 4.11: Constraints on σ_8, w and Ω_m using DES ($w(\theta)$ x $\gamma_t(\theta)$) in combination with Planck (solid purple) and DES in combination with Planck plus BAO, SN Ia and H_0 measurements (dashed red). Also shown are the constraints from Planck only (filled blue) and BAO, SN Ia and H_0 measurements only (filled yellow).

$w \approx -1.5$ and $\sigma_8 \approx 1$. This is in part because DES provides modest constraints on H_0 which help break the degeneracy between h and Ω_m in the CMB. In addition, the Planck dataset prefers higher values of σ_8 and h than the DES data, such that in combination, the two probes carve out a smaller area in parameter space. This produces strong constraints on w when the two datasets are combined. In combination with Planck, we find that $\Omega_m = 0.32 \pm 0.02$, $\sigma_8 = 0.88 \pm 0.03$ and $w = -1.15 \pm 0.09$.

Fig. 4.11 shows the result of combining our measurements with additional data sets beyond the CMB. The other probes that we consider are BAO measurements from 6dF (Beutler et al., 2011), BOSS (Anderson et al., 2014; Ross et al. , 2015), Supernova type Ia measurements (Betoule et al. , 2014) and direct measurements of H_0 (Efstathiou , 2014). These data sets alone give constraints of $\Omega_m = 0.33 \pm 0.02$ and $w = -1.07 \pm 0.06$ and no constraint on σ_8 (the posterior distribution on σ_8 is fully informed by the prior). Combining these data sets with DES and the CMB gives an improvement in precision and strengthens our results to $\Omega_m = 0.31 \pm 0.01$ and $\sigma_8 = 0.86 \pm 0.02$ and $w = -1.09 \pm 0.05$.

4.6.2 COMPARISON WITH DES COSMIC SHEAR

The Dark Energy Survey Collaboration (2015) measured the 2-point shear correlations, for the same DES-SV area and source catalogs. The best fitting cosmological parameters in that work were $\sigma_8 = 0.81^{+0.16}_{-0.26}$ and $\Omega_m = 0.36^{+0.09}_{-0.21}$. Figs. 4.4 and 4.5 show the constraints from the analysis presented in this work on those parameters together with constraints from the shear 2-point correlations for the ΛCDM and wCDM models, respectively. There is very good agreement between the two analyses and a similar degeneracy direction in the $\Omega_m - \sigma_8$ plane as well.

The shape of the contours for the two methods in Fig. 4.4 is somewhat different, with the cosmic shear contours being more elongated. We find that the

slope α in the derived parameter $S_8 \equiv \sigma_8(\Omega_m/0.3)^\alpha$ is 0.16 for $w(\theta)$ and $\gamma_t(\theta)$ instead of 0.478 for cosmic shear. In part because the covariance between Ω_m and σ_8 is weaker, the constraints on each parameter are slightly stronger for the $w(\theta)$ and $\gamma_t(\theta)$ case. The results in this analysis are less sensitive to errors in the lensing shear and redshift distribution of source galaxies since these do not impact $w(\theta)$ at all, and additive errors in the shear cancel out of $\gamma_t(\theta)$ at lowest order. On the other hand, cosmic shear measurements are unaffected by errors in the galaxy biasing model and systematic errors in the measurement of galaxy clustering. Furthermore, the derived parameter S_8 is better constrained by DES cosmic shear. While there is significant complementarity in the two measurements, they are also correlated because of the shared source galaxies. The combination of all three 2-point functions taking into account covariances is an important next step in the cosmological analysis of DES.

Cosmic shear measurements obtained from CFHTLenS (Heymans et al. , 2013) constrain the combination of $S_8 = \sigma_8(\Omega_m/0.3)^\alpha$, where $\alpha = 0.46$, to be $S_8 = 0.774^{+0.032}_{-0.041}$ as their primary result. Again, the directionality of the parameter degeneracy between σ_8 and Ω_m is slightly stronger than for our joint probes analysis, but the results appear consistent.

4.6.3 COMPARISON WITH THE LITERATURE

A number of previous papers have considered the combination of $w(\theta)$ and $\gamma_t(\theta)$ as probes of cosmology. Mandelbaum et al. (2013) perform an analysis with SDSS DR7 using luminous red galaxies as the lenses and derive comparable constraints. With some cosmological parameters fixed, Mandelbaum et al. (2013) used a combination of three lensing and angular clustering measurements in the redshift range $0< z <0.5$ to obtain $\sigma_8 = 0.76 \pm 0.08$ and $\Omega_m = 0.27^{+0.04}_{-0.03}$. Several details of our analysis differ from Mandelbaum et al. (2013), but the broad approach of employing a quasilinear analysis on large scales is similar and the results are consistent.

Cacciato et al. (2013) also measure the tangential shear and angular clustering from SDSS DR7 data, but differ in that they include small scale clustering and consider a subset of the galaxy samples used by Mandelbaum et al. (2013). They adopt a halo model approach which allows them to extend their analysis to much smaller scales than Mandelbaum et al. (2013), at the expense of requiring additional free parameters and model ingredients that are calibrated with simulations. With this small scale approach, Cacciato et al. (2013) obtain $\Omega_m = 0.278^{+0.023}_{-0.026}$ and $\sigma_8 = 0.763^{+0.064}_{-0.049}$, again consistent with our derived constraints.

Similarly, More et al. (2015) use a halo model approach to calculate the joint likelihood using galaxy clustering, galaxy-galaxy lensing and galaxy abundance

for the CMASS sample observed in BOSS using CFHTLenS sources. They report that $\Omega_m = 0.31 \pm 0.02$ and $\sigma_8 = 0.79 \pm 0.04$. Applying an HOD model motivates the inclusion of small scale information in their cosmology fits. In terms of number density and typical halo mass, the CMASS galaxies used by More et al. (2015) are closer to our redMaGiC sample than the LRGs in Mandelbaum et al. (2013), but they all derive consistent cosmological constraints.

4.7 CONCLUSIONS

In this chapter we have presented cosmological constraints from the combination of large-scale structure and weak gravitational lensing in the Dark Energy Survey. Using a contiguous patch of 139 sq. deg. from the Science Verification period of observations, we have placed constraints on the matter density and the amplitude of fluctuations in the Universe as $\Omega_m = 0.31 \pm 0.09$ and $\sigma_8 = 0.74 \pm 0.13$, respectively. We also present joint constraints with CMB measurements from Planck, and additional low-redshift datasets. When allowing for a dark energy equation of state parameter w different to the ΛCDM value of -1, we find DES data improve the constraints on σ_8 as well as w. We leave a full tomographic analysis with multiple lens bins and a joint analysis with cosmic shear for future DES releases.

We have assessed the robustness of our results with respect to several variations in the choice of data vector, modelling and treatment of systematics. In particular, the results are stable under the use of two different shear catalogs, four different photo-z codes and two different estimators of the lensing signal. They also show consistency with the fiducial results when using a different lens bin, a different selection of angular scales or when adding a nonlinear galaxy bias parameter.

The DES-SV region comprises only $\sim$3% of the eventual survey coverage, and we expect to greatly improve on our constraining power with future data releases. For now, the analysis presented in this chapter is complementary to and provides a useful consistency check with the analysis of the shear 2-point function presented in The Dark Energy Survey Collaboration (2015). These analyses validate the robust modelling of systematic errors and galaxy bias, as well as the exhaustive testing of the shear pipeline, photo-z estimation and the redMaGiC galaxy sample selection in the Dark Energy Survey.

Part IV

Cosmic voids and void lensing

Chapter 5

COSMIC VOIDS AND VOID LENSING IN THE DES-SV DATA

5.1 INTRODUCTION

Cosmic voids are low-density regions in space surrounded by a network of dark matter halos and the galaxies that populate them. Given their intrinsic low-density environment, voids are only weakly affected by complicated non-linear gravitational effects which have a strong impact in crowded environments such as galaxy clusters. This simplicity makes it possible to constrain cosmological parameters with voids (Betancort-Rijo et al., 2009; Lavaux & Wandelt, 2010; Sutter et al., 2014b; Kitaura et al., 2015; Hamaus et al., 2016; Mao et al., 2016; Sahlén, Zubeldía & Silk, 2016). Furthermore, the unique low-density environments of voids make possible probes of the nature of dark energy, alternate theories of gravity (Lee & Park, 2009; Bos et al., 2012; Spolyar, Sahlén & Silk, 2013; Cai, Padilla & Li, 2015; Barreira et al., 2015), and primordial non-Gaussianity (Song & Lee, 2009).

A number of different void finding algorithms exist in the literature: Voronoi tesselation and watershed methods (Platen, Van De Weygaert & Jones, 2007; Neyrinck, 2008; Lavaux & Wandelt, 2012; Sutter et al., 2012; Nadathur et al., 2015), growth of spherical underdensities (Hoyle & Vogeley, 2002; Colberg et al., 2005; Padilla, Ceccarelli & Lambas, 2005; Ceccarelli et al., 2006; Li, 2011), hybrid methods (Jennings, Li & Hu, 2013), 2D projections (Clampitt & Jain, 2015), dynamical criteria (Elyiv et al., 2015), and Delaunay Triangulation (Zhao et al., 2015), among other methods (Colberg et al., 2008). Most void finders currently applied to data use galaxies with spectroscopic redshifts to define voids. However, when using far less precise photometric redshifts (photo-z's), the void-finding process needs to be revisited to overcome the smearing in the line-of-sight position

of tracer galaxies.

Spectroscopic surveys like 2dF (Colless et al., 2001), VVDS (Le Fèvre et al., 2005), WiggleZ (Drinkwater et al., 2010) or BOSS (Dawson et al., 2013) provide 3D information of the galaxy distribution, but they are expensive in terms of time, and may suffer from selection effects, incompleteness and limited depth. In contrast, photometric surveys such as SDSS (York et al., 2000), PanSTARRS (Kaiser, Tonry & Luppino, 2000), KiDS (de Jong et al., 2013) or LSST (Tyson et al., 2003) are more efficient and nearly unaffected by selection bias, more complete and deeper, but do not provide complete 3D information of the galaxy distribution due to their limited resolution in the galaxy line-of-sight positions, obtained by measuring the photo-z of each galaxy from the fluxes measured through a set of broadband filters.

A few void catalogs exist that use photometric redshift tracers (Granett, Neyrinck & Szapudi, 2008). Many voids about the size of the photo-z error or smaller will not be found at all; in other cases, spurious, or *Poisson*, voids will appear in the sample due to photo-z scatter. For the larger voids in the sample, those with sizes much larger than the photo-z error, the photo-z scatter should not affect the void sample substantially. However, these huge voids are very few due to the rapidly falling size distribution of cosmic voids in the universe. In any case, it should also be possible to find voids smaller than the photo-z scatter, since the latter acts to smooth out the density field, but retains the topology of the large-scale structure to some extent. Therefore, by designing a void finding algorithm specifically for photometric redshift surveys, the purity and completeness of the resulting void sample can be improved.

Qualitatively, our void finding method can be understood with an analogy to galaxy clustering measurements. In that case, the ideal scenario is to measure the 3D correlation function of galaxies when spectroscopic redshifts are available. However, for photometric survey data sets, one usually avoids computing the 3D correlation function of galaxies because of the photo-z dispersion affecting the line-of-sight component. The standard approach is therefore to split galaxies into tomographic photometric redshift bins, and compute the 2D angular correlation function in the projection of each of these line-of-sight bins. The photometric redshift errors make the actual size of the redshift bins to be effectively comparable or larger than the photo-z scatter (see for instance Crocce et al. 2011). Then, in order to minimize the noise in the measurement, the optimal approach is to set the width of the redshift bins to be comparable or larger than the photo-z scatter. Finally, one measures the angular clustering in each of these redshift bins, and hence the evolution of clustering with redshift. In this work we present a void finder that follows the same approach: finding voids in the angular projection of the galaxy distribution in redshift slices that are broader than the photo-z disper-

sion, and then combining the slices to get the most of the line-of-sight information in the data.

Before applying the algorithm to the DES Science Verification (DES-SV) data set, we use simulations with mock spectroscopic and realistic photometric redshifts to validate the method, running the void finder in both cases and studying the differences among the void catalogs coming from the corresponding projected slices. Once the DES-SV void catalog is defined, we measure the weak gravitational lensing signal around voids and confirm the voids are also empty in the dark matter.

The plan of the chapter is as follows. In Sec. 2 we describe the Dark Energy Survey Science Verification data used in this chapter, together with the simulations used to test the validity of the finder. Section 3 presents the 2D angular void finder algorithm and some simulation tests comparing the algorithm output when using spectroscopic and photometric redshifts for the tracer galaxies. Then, in Sec. 4 we apply the algorithm to DES-SV data and discuss the choice of redshift slices and the way we deal with survey edge effects. Finally, in Sec. 5 we use the final DES-SV void catalog to measure the weak gravitational lensing around voids and we discuss our results and conclusions in Sec. 6.

5.2 DATA AND SIMULATIONS

The Dark Energy Survey (DES, Flaugher 2005; Flaugher et al. 2015; Dark Energy Survey Collaboration et al. 2016) is a photometric redshift survey that will cover about one eighth of the sky (5000 sq. deg.) to a depth of $i_{AB} < 24$, imaging about 300 million galaxies in 5 broadband filters ($grizY$) up to redshift $z = 1.4$. The DES camera (DECam, Flaugher et al. 2015) includes sixty-two 2048x4096 science CCDs, four 2048x2048 guider CCDs, and eight 2048x2048 focus and alignment chips, for a total of 570 megapixels. In this chapter we use 139 sq. deg. of data from the Science Verification (SV) period of observations (Diehl et al., 2014), which provided science-quality data at close to the nominal depth of the survey.

In a photometric redshift survey, such as DES, the photo-z's of tracer galaxies will impact the identification of voids with sizes comparable to the photo-z scatter σ_z, in a way that renders some voids smeared and undetected. For DES main galaxies, this is a problem since $\sigma_z \simeq 0.1$ (Sánchez et al., 2014), corresponding to ~ 220 Mpc/h at $z = 0.6$, and typical voids have a comoving size of about 10-100 Mpc/h. However, we do not need to use all DES galaxies as void tracers. Instead, we can restrict ourselves to the Luminous Red Galaxies (LRGs) in the sample, which are still good tracers of the large-scale structure and have much better photo-z resolution.

5.2.1 VOID TRACER GALAXIES: THE REDMAGIC CATALOG

The DES-SV redMaGiC catalog (Rozo et al., 2015) presents excellent photo-z performance: redMaGiC photometric redshifts are nearly unbiased, with median bias $(z_{\rm spec}-z_{\rm phot}) \approx 0.5\%$, a scatter $\sigma_z/(1+z) \approx 1.70\%$, and a $\approx 1.4\%$ 5σ redshift outlier rate. That scatter corresponds to a redshift resolution of ~ 50 Mpc/h at $z = 0.6$, a substantial improvement over DES main galaxies. Next we summarize the redMaGiC selection algorithm, but we refer the reader to Rozo et al. (2015) for further details.

The red-sequence Matched-filter Galaxy Catalog (redMaGiC, Rozo et al. 2015) is a catalog of photometrically selected luminous red galaxies (LRGs). We use the terms redMaGiC galaxies and LRG interchangeably. Specifically, redMaGiC uses the redMaPPer-calibrated model for the color of red-sequence galaxies as a function of magnitude and redshift (Rykoff et al., 2014). This model is used to find the best fit photometric redshifts for all galaxies under the assumption that they are red-sequence members, and the χ^2 goodness-of-fit of the model is then computed. For each redshift slice, all galaxies fainter than some minimum luminosity threshold $L_{\rm min}$ are rejected. In addition, redMaGiC applies a cut $\chi^2 \leq \chi^2_{\rm max}$, where the cut $\chi^2_{\rm max}$ as a function of redshift is chosen to ensure that the resulting galaxy sample has a constant space density $\bar{n}$. In this work, we set $\bar{n} = 10^{-3}h^3{\rm Mpc}^{-3}$ with ΛCDM cosmological parameters $\Omega_\Lambda = 0.7$, $h_0 = 100$, and redMaGiC galaxies are selected in the redshift range $0.2 < z < 0.8$. We expect the redMaGiC galaxy selection to be only marginally sensitive to the cosmological parameters assumed (see Rozo et al. 2015 for details). The luminosity cut is $L \geq L_*(z)/2$, where the value of $L_*(z)$ at $z = 0.1$ is set to match the redMaPPer definition for SDSS (Rykoff et al., 2014), and the redshift evolution for $L_*(z)$ is that predicted using a simple passive evolution starburst model at $z = 3$ (Bruzual & Charlot, 2003).

We use the redMaGiC sample because of the exquisite photometric redshift performance of the redMaGiC galaxy catalog. Also, because void properties depend on the tracer sample used, the constant comoving density of redMaGiC tracers helps in assuring the resulting voids have similar properties. For example, the dark matter profile (Sutter et al., 2014a) and void bias (Chan, Hamaus & Desjacques, 2014; Clampitt, Jain & Sánchez, 2016; Pollina et al., 2016) have been shown to depend on the tracer density or tracer bias used to define voids.

Aside from the data catalog presented above, in this work we also use ΛCDM simulations that mimic the properties of the DES-SV redMaGiC data set. The mock galaxy catalog is the Buzzard-v1.0 from the Blind Cosmology Challenge (BCC) simulation suite, produced for DES (Wechsler et al, in preparation). These catalogs have previously been used for several DES studies (see e.g. Chang et al.

2015; Leistedt et al. 2015; Becker et al. 2015; Clampitt et al. 2016; Kwan et al. 2016). The underlying N-body simulation is based on three cosmological boxes, a 1050 Mpc/h box with 1400^3 particles, a 2600 Mpc/h box with 2048^3 particles and a 4000 Mpc/h box with 2048^3 particles, which are combined along the line of sight producing a light cone reaching DES full depth. These boxes were run with LGadget-2 (Springel, 2005) and used 2LPTic initial conditions (Crocce, Pueblas & Scoccimarro, 2006) with linear power spectra generated with CAMB (Lewis & Bridle, 2002). ROCKSTAR (Behroozi, Wechsler & Wu, 2013) was utilized to find halos in the N-body volumes. The ADDGALS algorithm (Wechsler 2004, Busha et al. 2013, Wechsler et al, in preparation) is used to populate the dark matter simulations with galaxies as a function of luminosity and color. ADDGALS uses the relationship between local dark matter density and galaxy luminosity, to populate galaxies directly onto particles in the low-resolution simulations. This relationship is tuned to reproduce the galaxy–halo connection in a higher resolution tuning simulation, in which galaxies are assigned using subhalo abundance matching (e.g. Conroy, Wechsler & Kravtsov, 2006; Reddick et al., 2013), in this case matching galaxy luminosity to peak circular velocity. Finally, each galaxy is assigned a color by using the color-density relationship measured in the SDSS (Aihara et al., 2011) and evolved to match higher redshift observations. The redMaGiC algorithm has been run on the simulation in a similar way as it is run on the DES data. This produces a simulated sample with the same galaxy selection and photometric redshift performance as the DES-SV redMaGiC catalog but gives us access to the true redshifts of the galaxies in the sample, a fact that we will use to test the void finder presented in this work.

5.2.2 LENSING SOURCE CATALOG

The catalog of galaxy shapes used in the lensing measurement of this work is the `ngmix`[1] catalog presented in Jarvis et al. (2015). `ngmix` is a shear pipeline which produces model fitting shape measurements, and that was applied to a large subset of DES-SV galaxies, meeting the requirements of an extensive set of null and systematics tests in Jarvis et al. (2015). The photometric redshifts of the galaxies in the `ngmix` shear catalog were studied in detail in Bonnett et al. (2015), using 4 different photo-z codes. In this work we use the SkyNet photo-z method, which demonstrated excellent performance in that comparison.

[1] https://github.com/esheldon/ngmix

5.3 PHOTO-Z VOID FINDER ALGORITHM

In this Section we present a new void finder designed specifically to work on photometric surveys. We explain the algorithm and test its performance on simulations, providing validation for the results shown later in the chapter.

5.3.1 VOID FINDER ALGORITHM

The void finder works by projecting galaxies in redshift slices and finding underdensities in the 2D angular distribution of galaxies in the given slices. If the line-of-sight width of the projected slice is sufficiently large, at least about twice the photo-z resolution, then most galaxies will still be assigned to the correct slice. Since the finder works by projecting all galaxies within a given slice onto a 2D surface, the line-of-sight position within the slice does not affect the results.

The void finder of Clampitt & Jain (2015) also begins by treating each slice in isolation, but has the disadvantage that voids are required to be completely empty of galaxies near the center. Thus, photo-z scatter, which moves a single galaxy between slices, can inappropriately break up a single large void into several smaller voids, or even result in no void being detected at all. To overcome this problem, we smooth the 2D projected galaxy density field in each slice and then voids are found from minima of the smoothed density field. This means a few galaxies moving between different slices will not greatly affect the resulting set of voids, as will be demonstrated in Sec. 5.3.2.

In detail, the void finding algorithm involves the following steps:

1. We select the galaxies from a redshift slice of thickness $2s_v$ (we define s_v to be half the slice thickness) and we project them into a HEALpix map (Gorski et al., 2005), with a resolution of $N_{\rm side} = 512$ representing an angular resolution of 0.1 deg. and a physical resolution of 1.5 Mpc/h at $z = 0.3$ (3 Mpc/h at $z = 0.6$).

2. We compute the mean density in the map corresponding to the given redshift slice, $\bar{n}_{2d}$, and convert the galaxy map to a density contrast map as $\delta = n_{2d}/\bar{n}_{2d} - 1$, where n_{2d} is the galaxy map.

3. Then we smooth the density contrast map with a Gaussian filter of comoving scale $\sigma_s = 10$ Mpc/h.

4. We take this smoothed contrast map and consider only the most underdense pixels (with $\delta < \delta_m = -0.3$) as potential void centers. We define the most underdense pixel in the map as the first void center.

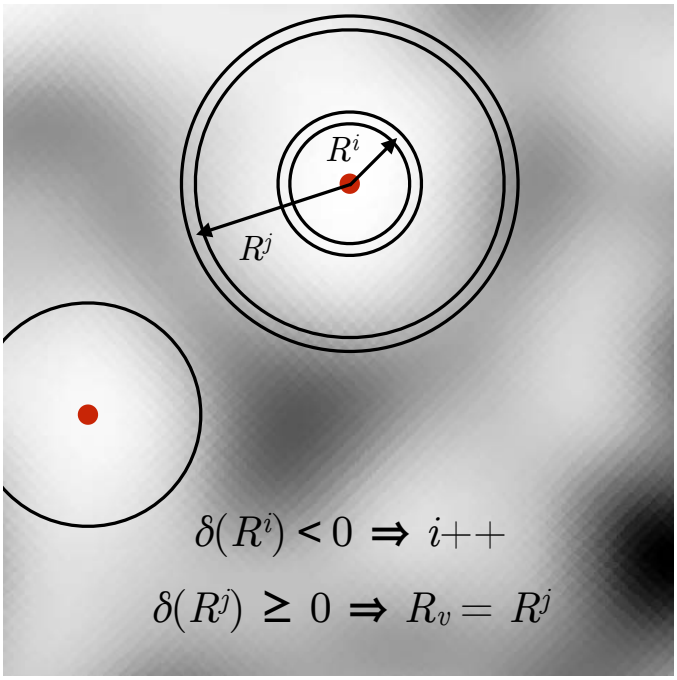

Figure 5.1: Graphical description of the void-finding algorithm presented in this chapter. The background gray-scaled field is the smoothed galaxy field ($\sigma = 10$ Mpc/h) in a redshift slice used by the void-finder. The two solid (red) dots show two void centers. For the upper void, we show a circular shell or radius R^i. Since the density contrast $\delta(R^i) < 0$, the algorithm checks larger shells, up to radius R^j such that $\delta(R^j) \geq 0$. The void radius is then defined as $R_v = R^j$.

5. Next we start defining circular shells of increasing radius around that center, stopping when the mean density within the slice ($\delta = 0$) is reached. That is, starting with a shell of radius R_v^i, we measure the average galaxy density in the shell $\delta(R_v^i)$, and if the density is negative we check the next larger shell $\delta(R_v^{i+1})$, where the increment between shells is 1 Mpc/h in radius. For some shell R_v^j the density contrast reaches zero, $\delta(R_v^j) \geq 0$, and at that point the void radius is defined as $R_v = R_v^j$ (see Fig. 5.1 for a graphical explanation).

6. Then all pixels contained in this void are removed from the list of potential void centers, preventing any of these pixels to become the center of any other void. From the remaining pixels that satisfy $\delta < \delta_m = -0.3$, we define the next most underdense pixel as the second void center. The process is repeated until all pixels with $\delta < \delta_m = -0.3$ have been assigned to a void.

Beyond the dependency on the line-of-sight size of the projected slice in which the finder is executed, studied in more detail later in this section, the void catalog produced by this algorithm depends on two parameters: the smoothing scale, σ_s, and the maximum density contrast of a pixel to become a void center, δ_m. The smoothing scale ($\sigma_s = 10$ Mpc/h) is chosen to be about half the radius of the

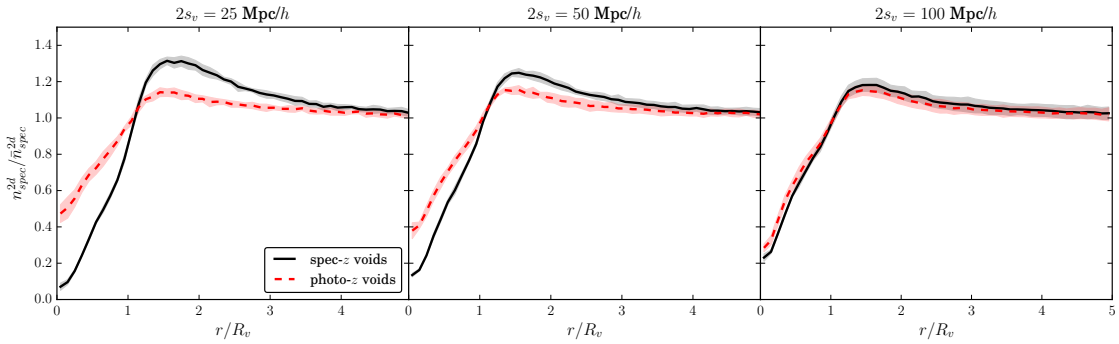

Figure 5.2: (*left panel*): Comparison of 2D spectroscopic galaxy density profiles of voids found in the simulations using galaxy spectroscopic redshifts (solid line) or photometric redshifts (dotted, red). The shaded regions show the corresponding error bars computed as the standard deviation among all the stacked voids. The projected 2D slice width is 25 Mpc/h(comoving distance), a scale corresponding to $\sim 1/2$ the photometric redshift scatter. For this thin slice, the galaxy density profile is damped significantly by photo-z scatter, making the galaxy profile of photo-z defined voids more shallow. (*center panel*): The same, but for a thicker slice of width 50 Mpc/h, comparable to the photo-z scatter. (*right panel*): The same, but for a projected slice of width 100 Mpc/h, twice the size of the typical photo-z scatter. In this case there is a good match between the profiles of spec-z and photo-z selected voids. For such a thick slice, the fraction of galaxies that are placed in the incorrect slice due to photometric redshift scatter is smaller, allowing accurate void identification from the smoothed galaxy field.

smallest voids we can access in our data sample (because of photo-z smearing), and increasing it would erase the structure leading to some of these smallest voids, leaving the large voids intact. On the other hand, the most significant voids found by the algorithm, the deepest ones, are independent of the choice $\delta_m = -0.3$ since their void center pixel is more underdense than that. By changing the value of δ_m we are only affecting the shallower voids of the sample. The impact of the δ_m choice is studied in Appendix D. Also, voids found by this algorithm can overlap or even enclose one another, but just in the case where a subvoid is deeper than the bigger void enclosing it.

The process detailed above will produce a list of voids for a given redshift slice. Before describing how various slices are combined to obtain the full void catalog, we first study the performance of the single slice results in simulations.

5.3.2 PERFORMANCE ON SIMULATIONS

In order to validate the performance of the algorithm we use the simulations, where we have both spectroscopic and photometric redshift for void tracer galaxies, and we compare the voids found by the algorithm in spec-z and photo-z space. In particular, we run the void finding algorithm twice on each redshift slice: first using spectroscopic redshifts for selecting the galaxies that go into the slice and then using photometric redshifts that mimic the ones we have in real DES data.

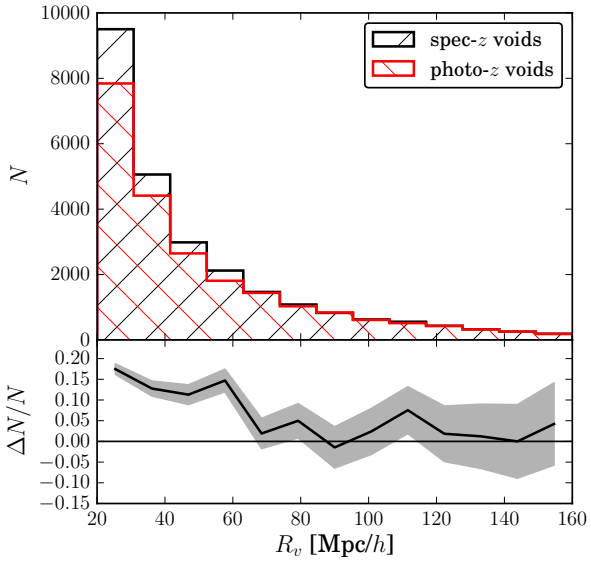

Figure 5.3: (*upper panel*): Void radius distribution for voids found in spec-z and photo-z simulated galaxy samples, for a slice thickness of $2s_v = 100$ Mpc/h. (*lower panel*): Relative difference between the distributions (with respect to the spectroscopic redshift case). Some voids with size smaller than the photo-z scatter ($\sigma_z \simeq 50$ Mpc/h) are smeared out due to photo-z scatter and not detected, resulting in a smaller number of voids relative to the spectroscopic case. For large voids this effect is not important and the two distributions agree within errors.

Once we have the spec-z and photo-z defined void catalogs, we measure the projected galaxy density profiles of the voids in them in radial annuli using the true redshifts. Figure 5.2 shows the resulting density profiles for both cases in different slice comoving thicknesses. As expected, the void finder performs poorly if the size of the projected slice is smaller or similar to the photo-z dispersion $\sigma_z \simeq 50$ Mpc/h. Therefore, the accuracy of the finder is a function of the thickness of the projected slice: for slice width ~ 2 times the size of the typical photometric redshift scatter, the difference between the average density profiles of voids found in spec-z and photo-z is not significant, being smaller than the standard deviation of the stacked void profiles.

Figure 5.2 shows that voids found by the algorithm in photo-z space can indeed have very similar density profiles as voids found in spec-z space. However, it is also important to know the relative number of voids found in the two cases. Photometric redshifts produce a smearing in the line-of-sight position of tracers that can actually erase some of the structure, especially on scales comparable to the size of the photo-z scatter or smaller. That will have the consequence of some small voids not being detected in the photo-z case. The voids of size larger than the photo-z scatter should be detected in both cases. Figure 5.3 shows the distribution of void radii in simulations for spec-z and photo-z samples. As expected, we find less voids in the photo-z case, with the difference being more important for small voids and becoming negligible for the voids substantially larger than the photo-z dispersion ($\sigma_z \simeq 50$ Mpc/h).

In addition to the comparison of the galaxy density profiles of voids, which is the most important test of the algorithm, Fig. 5.4 shows a visual comparison between the positions and radius of spec-z and photo-z defined voids in a random

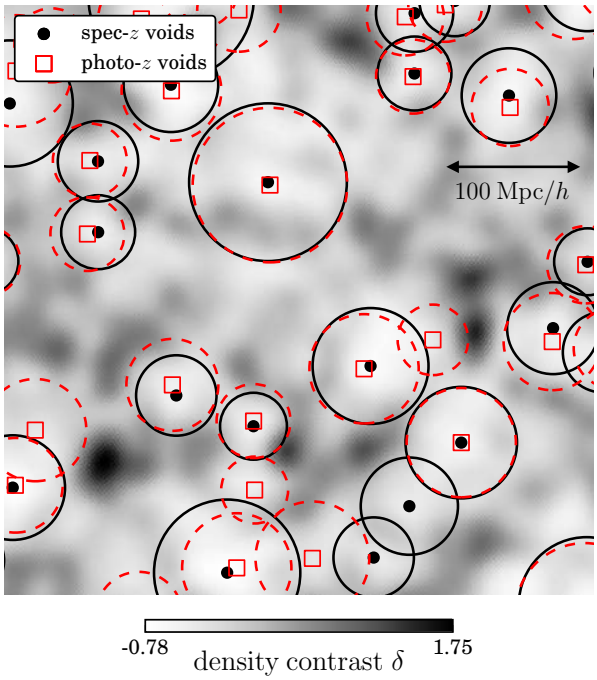

Figure 5.4: Comparison between voids found in spec-z (centers: solid black points; radius: solid circles) and photo-z (centers: open red squares; radius: red dashed circles) in the simulations for a slice of thickness $2s_v = 100$ Mpc/h. The background gray-scaled field is the smoothed galaxy field ($\sigma = 10$ Mpc/h) used by the void-finder. The correlation between spec-z and photo-z defined voids is clear: in many cases the void position and radius match almost exactly.

100 Mpc/h-thick slice of our simulations. The correlation between the two sets of voids is very clear, in both positions and radii. In some cases, especially for the biggest voids, the match between spec-z and photo-z voids is almost perfect. This is remarkable given the magnitude of the scatter in the line-of-sight direction being added by photometric redshifts.

5.4 DES-SV VOID CATALOG

In the previous Section we have presented a void finder that works by projecting galaxies into redshift slices (see Sec. 5.3.1 for a detailed description and parameters used in the algorithm). We have shown (Sec. 5.3.2) that as long as the thickness of the projected slice is large enough compared to the photo-z scatter, using photometric redshifts for the position of void tracers works nearly as well as using spectroscopic redshifts. Nevertheless, the algorithm will find some voids that are not likely to correspond to voids in the dark matter density field. Such false voids may be due to a number of effects: (i) at the survey edge or masked areas we have no information on galaxy positions, and (ii) duplicate voids may appear if slices overlap in redshift. In this Section we apply the algorithm to real DES-SV data, and present the way we deal with voids near the survey edge (Sec. 5.4.1) and the strategy we follow to get the most of the line-of-sight information in the data (Sec. 5.4.2). The properties of the final DES-SV void catalog are presented in Sec. 5.4.3.

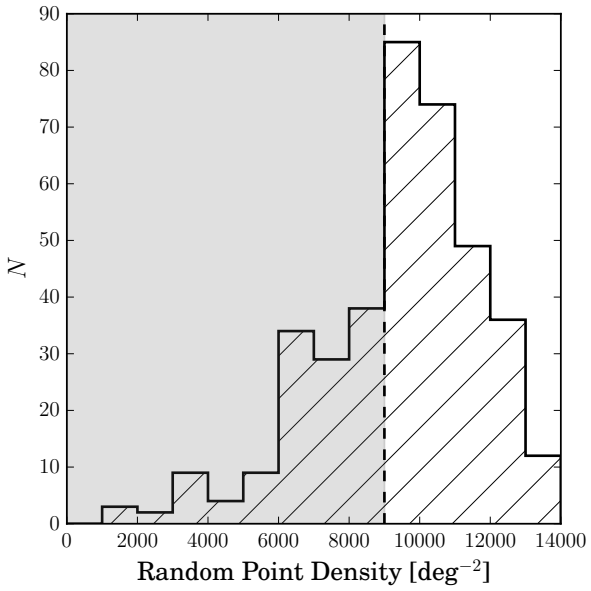

Figure 5.5: Distribution of random point density inside DES-SV voids, where the random points are distributed uniformly through the DES-SV area. The distribution shows roughly a Gaussian shape at high densities corresponding to voids inside the survey mask, and a low density tail corresponding to edge voids. We remove all voids with random point density less than 9000 points/deg^2 (shaded region), most of them near the survey edge. This cut removes 33% of the total number of voids.

5.4.1 VOIDS NEAR THE SURVEY EDGE

The assignment of each void's radius does not distinguish between voids that are fully contained within the survey and those that extend beyond it. The void radius may stretch beyond the edge of the survey, into areas which may or may not correspond to voids in the galaxy distribution. To remove such voids which extend far beyond the survey edge, we use the method of Clampitt & Jain (2015). A random point catalog drawn using the survey mask is generated, and for each void we calculate the density of random points inside R_v. The distribution of random points density inside voids is shown in Fig. 5.5, and it presents a Gaussian-like shape at high densities (peaked around 9500 points/deg^2 with $\sigma \simeq 2000$ points/deg^2), corresponding to voids centered in the survey mask, and a low density tail reaching almost zero density, which corresponds to edge voids. Due to the small size of the DES-SV patch used in this work, with an area of 139 sq. deg., and the size of some of the voids detected (a void with $R_v \sim 80$ Mpc/h would span more than 10 deg. in diameter at $z = 0.3$), we place a conservative cut and discard voids with random point density less than 9000 points/deg^2, which constitute 33% of the total number of voids.

5.4.2 LINE OF SIGHT SLICING STRATEGY

To obtain more information about the line-of-sight position of each void we oversample the volume with a number of different slice centers. In particular, first we slice the line-of-sight range of the survey, $0.2 < z < 0.8$, in equal slices of comoving thickness $2s_v = 100$ Mpc/h taking the upper redshift limit, $z = 0.8$,

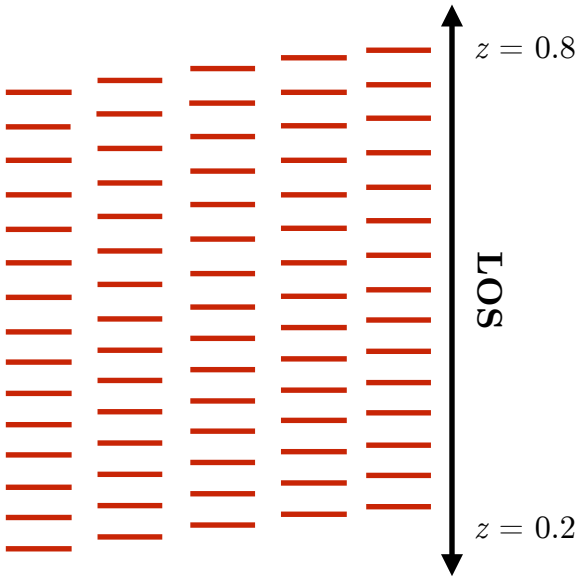

Figure 5.6: Graphical representation of the line-of-sight (LOS) slicing performed in this chapter. The black vertical arrow represents the redshift range, $0.2 < z < 0.8$, and the red horizontal bars represent the boundaries of the redshift slices in which the void finder is run. As the diagram shows, we oversample the line of sight with slices of thickness 100 Mpc/h every 20 Mpc/h. In Fig. 5.7 we show the way voids in adjacent slices are combined to form the final catalog.

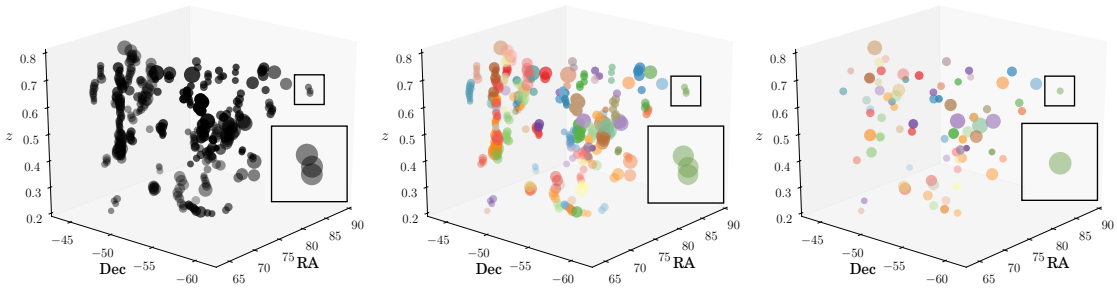

Figure 5.7: (*left panel*): 3D position of voids found in the slicing shown in Fig. 5.6. Each void candidate is shown as a sphere with size proportional to the void radius. Due to oversampling in the line of sight, slices overlap and duplicates of the same physical void are found in different slices, apparent in this plot as elongated structures in redshift. The inset square shows the case of a three void group. (*center panel*): Voids corresponding to the same physical underdensity are grouped together (as described in Sect. 5.4.2) and plotted with a common color. (*right panel*): The final void positions are computed as the median 3D position of the members of each group.

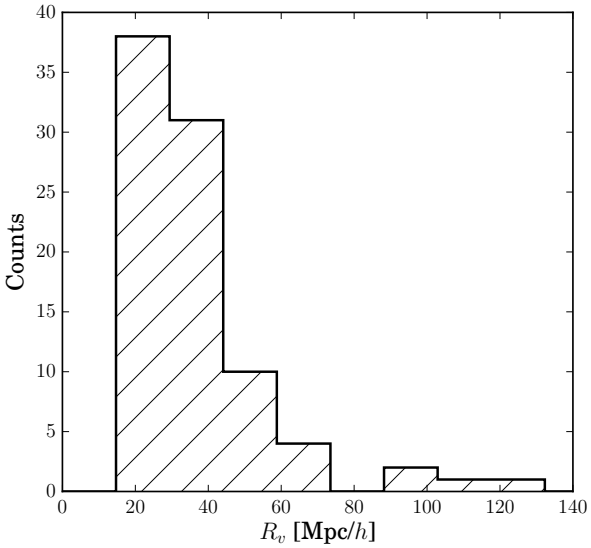

Figure 5.8: Distribution of comoving void radii of the final DES-SV void catalog used in this work, using slices of thickness 100 Mpc/h and after the cuts described in Sections 5.4.1 and 5.4.2.

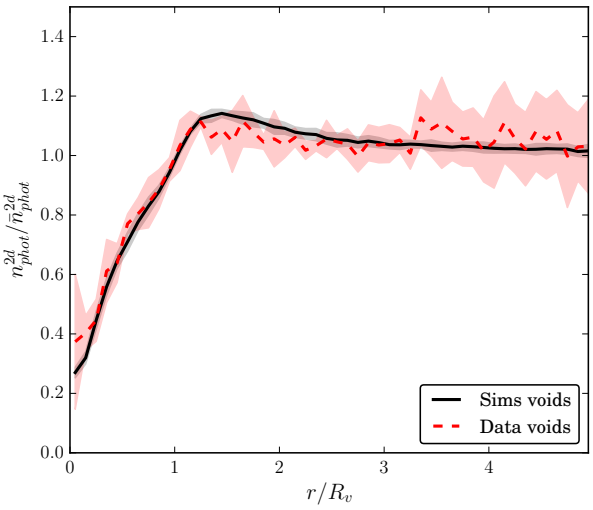

Figure 5.9: Comparison of 2D galaxy density profiles of voids found in DES-SV data and simulations, using galaxy photometric redshifts. The shaded regions show the corresponding error bars computed as the standard deviation among all the stacked voids.

as the upper limit of the furthest slice. Then we apply a shift to this slicing of 20 Mpc/h towards low redshift, and we repeat the process four times so that we have a slice of thickness 100 Mpc/h centered every 20 Mpc/h of the line-of-sight range in the data (see Fig. 5.6 for a graphical representation).

Since the volume has been oversampled with a number of different slice centers, sometimes the same physical void will be found in multiple slices, creating elongated void structures in the line of sight (left panel in Fig. 5.7). Each of these structures may actually correspond to one physical underdensity, or at least their void candidate members will have a consistent lensing profile since they are essentially at the same redshift and have very similar sizes. In order to remove the duplicate voids, and also to pick up the right void center in the line-of-sight direction, we need to group these void structures together. The groups are found by

joining voids in neighboring (and hence overlapping) slices that have a small angular separation between them. In particular, two voids with radii R_v^i and R_v^j and found in neighboring slices will become part of the same group if the angular distance between their centers is smaller than half the mean angular radii of the two voids: $\bar{R_v}/2 = (R_v^i + R_v^j)/4$. The groups are shown in the central panel in Fig. 5.7, and the right panel shows the final void catalog, without obvious elongated structures in the line of sight. This resulting void catalog is not very sensitive to the choice of $\bar{R_v}/2$: Increasing this minimum separation from $0.5\bar{R_v}$ to $0.6\bar{R_v}$ ($0.8\bar{R_v}$) results in removing 6% (10%) of the voids in the final catalog.

Once we have the void groups corresponding to those line-of-sight structures, we compute the 3D position of each group (RA, Dec and redshift) as the median position of the different void members of the group. The relative scatter in this determination inside each group (taken as the standard deviation of each quantity with respect to its mean value) is very small (less than 0.4% for RA and Dec and around 2% in redshift). The void radius is also computed as the median void radius of the different void members in each group, with a relative scatter around 14%. The final void candidates, after removal of duplications of potential physical underdensities due to the oversampled slicing, are shown in the right panel of Fig. 5.7. The effect of the LOS slicing strategy in the void lensing measurement is tested in Appendix E, where we show it helps reduce the noise but it does not affect the main outcomes from the measurement.

5.4.3 FINAL VOID CATALOG

Applying the void finding algorithm described in Sect. 5.3, using slices of 100 Mpc/h thickness, to the DES-SV redMaGiC catalog, and after making the cuts presented in Sections 5.4.1 and 5.4.2, we find a total of 87 voids in the 139 sq. deg. of survey area. These voids are identified in the redshift range $0.2 < z < 0.8$, and they have comoving sizes ranging from $R_v = 18$ Mpc/h to $R_v = 120$ Mpc/h, with a mean void radius of $\bar{R_v} = 37$ Mpc/h. Figure 5.8 shows the full void radius distribution for the sample. The mean angular radius of voids in the sky is 1.5 degrees while their mean redshift is $\bar{z} = 0.57$.

Figure 5.9 shows the 2D galaxy density profiles of voids found in the DES-SV data and in simulations, using galaxy photometric redshifts. The agreement between data and simulations is good, and so is the agreement between the simulation profiles measured with photometric (Fig. 5.9) and spectroscopic redshifts (*right panel* of Fig. 5.2).

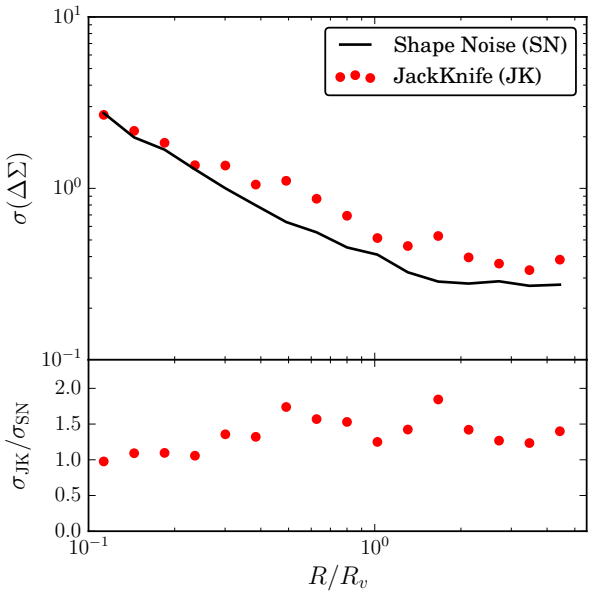

Figure 5.10: *(Upper panel):* Variance in the stacked weak lensing measurement of voids in DES-SV data, in bins of R/R_v, as estimated from jackknife (JK) resampling and lensing shape noise, the two techniques described in Sect. 5.5.2. *(Lower panel):* Ratio of the two error estimations in the upper panel. The two agree well on small scales (which are shape noise dominated) and differ significantly at medium to large scales since the jackknife includes other sources of variance in addition to shape noise.

5.5 VOID LENSING

Using the void catalog defined in the previous section we now focus on the lensing measurement around voids. This represents a key result, since a significant lensing signal around voids proves them to be underdense in the matter field, this way demonstrating the void catalog is primarily composed of real underdensities rather than spurious detections, tracer density effects or any systematics in the data.

In this section we present the details of the lensing measurement and covariance, the results for the tangential and cross components of that measurement and their significance, and the fit of the tangential component to a void model widely used in the literature.

5.5.1 MEASUREMENT

Assuming an axisymmetric density profile, the stacked excess surface mass density $\Delta\Sigma$ is related to the tangential shear γ_t of source galaxies by

$$\Delta\Sigma(R/R_v) = \Sigma_{\rm crit}\gamma_t(R/R_v)\,, \tag{5.1}$$

where the proportionality factor describing the lensing strength is

$$\Sigma_{\rm crit}(z_{\rm L}, z_{\rm s}) = \frac{c^2}{4\pi G}\frac{D_A(z_{\rm s})(1+z_{\rm L})^{-2}}{D_A(z_{\rm L})D_A(z_{\rm L}, z_{\rm s})}\,, \tag{5.2}$$

with $\Sigma_{\rm crit}^{-1}(z_{\rm L}, z_{\rm s}) = 0$ for $z_{\rm s} < z_{\rm L}$, where $z_{\rm L}$ and $z_{\rm s}$ are the lens and source galaxy redshifts, respectively. Note both the use of comoving units and that we need to assume a certain cosmology (flat ΛCDM with $\Omega_m = 0.3$) when calculating the angular diameter distances D_A in $\Sigma_{\rm crit}$. Our lensing projected surface density estimator is therefore given by

$$\Delta\Sigma_k(R/R_v; z_{\rm L}) = \frac{\sum_j \left[w_j \gamma_{k,j}(R/R_v) \Sigma_{{\rm crit},j}(z_{\rm L}, z_{\rm s}) \right]}{\sum_j w_j} \tag{5.3}$$

where k denotes the two possible components of the shear (tangential and cross), the summation $\sum_j$ runs over all the source galaxies in the radial bin R/R_v, around every void position, and the optimal weight for the j-th galaxy is given by (Sheldon et al., 2004):

$$w_j = \frac{[\Sigma_{{\rm crit},j}^{-1}(z_{\rm L}, z_{\rm s})]^2}{\sigma_{\rm shape}^2 + \sigma_{{\rm m},j}^2}\,. \tag{5.4}$$

Here $\sigma_{\rm shape}$ is the intrinsic shape noise for each source galaxy, and $\sigma_{{\rm m},j}$ is the shape measurement error. In Sec. 5.5.5 we relate the differential surface density $\Delta\Sigma$ to the 3D void profile $\rho_{\rm v}$.

Note that since the projected void radius R_v ranges from 20 to more than 100 Mpc/h, we stack the measured shear profiles in units of the void radius, R/R_v. Stacking the profiles in physical distance would smooth out the stacked void density profiles and hence some of the signal would be lost.

5.5.2 COVARIANCE

In order to estimate the covariance for the $\Delta\Sigma(R)$ measurements in this work we combine two different approaches: we rely on the jackknife (JK) method to estimate the signal variance while we estimate the off-diagonal shape of the covariance from the lensing shape noise of the measurement (Melchior et al., 2014). The main reason for that combination is the limitation in the JK technique due to the small number of voids (~ 100) in our catalog, yielding very noisy off-diagonal correlations. However, we can obtain smooth shape-noise-only covariances by applying any number of random rotations to the ellipticities of source galaxies. Next we explain the precise combination of the two approaches.

Due to the small number of voids in the DES-SV catalog, we perform a void-by-void jackknife: we carry out the measurement multiple times with each void omitted in turn to make as many jackknife realizations as voids we have in the sample (N). Then, the variance of the measurement (Norberg et al., 2009) is

given by

$$\sigma^2_{\rm JK}(\Delta\Sigma_i) = \frac{(N-1)}{N} \times \sum_{\rm JK-}^{N} \Big[(\Delta\Sigma_i)^{{\rm JK}-k} - \overline{\Delta\Sigma_i}\Big]^2 \tag{5.5}$$

where the mean value is

$$\overline{\Delta\Sigma_i} = \frac{1}{N} \sum_{\rm JK-}^{N} (\Delta\Sigma_i)^{{\rm JK}-k}\,, \tag{5.6}$$

and $(\Delta\Sigma_i)^{{\rm JK}-k}$ denotes the measurement from the k-th JK realization and the i-th spatial bin.

The shape noise (SN) covariance of the measurement is estimated by randomly rotating the orientation of each source galaxy ellipticity many times ($N_{\rm SN} = 300$ in this analysis) and repeating the $\Delta\Sigma$ lensing measurement each time. Then the covariance is estimated as:

$$\begin{aligned}{\rm Cov}_{\rm SN}[\Delta\Sigma_i,\Delta\Sigma_j] = &\frac{1}{N_{\rm SN}} \\ &\times \sum_{\rm SN-}^{N_{\rm SN}} \Big[(\Delta\Sigma_i)^{{\rm SN}-k} - \overline{\Delta\Sigma_i}\Big]\Big[(\Delta\Sigma_j)^{{\rm SN}-k} - \overline{\Delta\Sigma_j}\Big]\end{aligned} \tag{5.7}$$

where the mean value is

$$\overline{\Delta\Sigma_i} = \frac{1}{N} \sum_{\rm SN-}^{N} (\Delta\Sigma_i)^{{\rm SN}-k}\,, \tag{5.8}$$

and $(\Delta\Sigma_i)^{{\rm SN}-k}$ denotes the measurement from the k-th shape noise (SN) realization and the i-th spatial bin.

Figure 5.10 shows a comparison of the measurement variance estimated from jackknife and shape noise, following the techniques described above. The errors coming from the two approaches agree well on the smallest scales, as expected since the small-scale regime is dominated by shape noise. However, at mid to large scales ($R \sim 0.28R_v$ and above) the JK errors get bigger than SN only, as they can trace other effects such as systematics in the data or sample variance. The shape noise calculation is, on the other hand, more adequate for off-diagonal elements of the covariance since it avoids the intrinsic noise limitation of the JK technique. Hence, in order to have a smooth covariance matrix with variance accurately estimated from JK, we follow the approach of fixing the shape of the covariance as given by the shape noise calculation, and renormalize it to the JK

estimates of the variance:

$$\mathrm{Cov}[\Delta\Sigma_i, \Delta\Sigma_j] = \mathrm{Corr}_{\mathrm{SN}}[\Delta\Sigma_i, \Delta\Sigma_j]\sigma_{\mathrm{JK}}(\Delta\Sigma_i)\sigma_{\mathrm{JK}}(\Delta\Sigma_j) \tag{5.9}$$

where $\mathrm{Corr}_{\mathrm{SN}}[\Delta\Sigma_i, \Delta\Sigma_j]$ is the shape noise correlation matrix (or reduced covariance) given by:

$$\mathrm{Corr}_{\mathrm{SN}}[\Delta\Sigma_i, \Delta\Sigma_j] = \frac{\mathrm{Cov}_{\mathrm{SN}}[\Delta\Sigma_i, \Delta\Sigma_j]}{\sigma_{\mathrm{SN}}(\Delta\Sigma_i)\sigma_{\mathrm{SN}}(\Delta\Sigma_j)} \tag{5.10}$$

The approach of renormalizing a smooth covariance to a JK estimated variance has been used before in the literature, for example by Crocce et al. (2016).

5.5.3 NULL TESTS: CROSS-COMPONENT AND *randomized* VOIDS

The cross-component of the measurement described in Sect. 5.5.1 is not produced by gravitational lensing and therefore is expected to vanish at first order. Similarly, the tangential component of the same measurement around *randomized* voids, which follow the size and redshift distribution of true voids but are randomly distributed in the survey area (Appendix F), is also expected to vanish. Figure 5.11 shows the cross-component of the stacked lensing measurement for true voids and the tangential component for *randomized* voids.

With $\mathrm{dof} = N_{\mathrm{bin}}$ as the number of R/R_v bins in the measurement and no model parameters, the null hypothesis χ^2 can be computed as

$$\chi^2_{\mathrm{null}} = \sum_{i,j} \Delta\Sigma_i \mathrm{Cov}^{-1}_{ij} \Delta\Sigma_j \tag{5.11}$$

where i, j correspond to radial bins in $\Delta\Sigma$ and Cov is the covariance matrix.

The cross-component of the measurement yields a $\chi^2_{\mathrm{null}}/\mathrm{dof} = 8.2/16$, and the tangential measurement around *randomized* voids, which are 10 times more numerous than true voids and whose production is described in greater detail in Appendix F, yields a $\chi^2_{\mathrm{null}}/\mathrm{dof} = 18.7/16$, both showing consistency with the null hypothesis.

5.5.4 TANGENTIAL SHEAR PROFILE

Figure 5.12 shows the measurement of the tangential component of the stacked lensing signal around voids. Assuming a non-central χ^2 distribution we can compute the signal-to-noise (S/N) of the measurement as

$$(S/N)^2 = \chi^2_{\mathrm{null}} - \mathrm{dof} = \sum_{i,j} \Delta\Sigma_i \mathrm{Cov}^{-1}_{ij} \Delta\Sigma_j - N_{\mathrm{bin}} \tag{5.12}$$

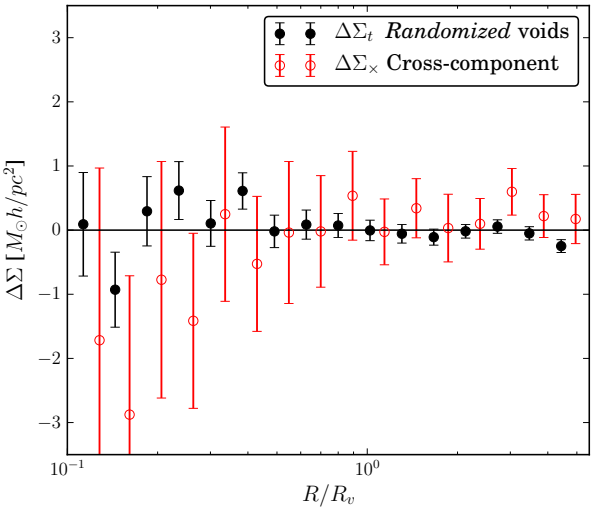

Figure 5.11: Cross-component of the DES-SV data stacked lensing measurement for true voids and tangential component for the lensing around *randomized* voids, in bins of R/R_v. Both measurements are compatible with the null hyposthesis with $\chi^2_{\rm null}/{\rm dof} = 8.2/16$ and $\chi^2_{\rm null}/{\rm dof} = 18.7/16$, respectively. The error using *randomized* voids is smaller since the measurement involves ~ 10 times more *randomized* voids.

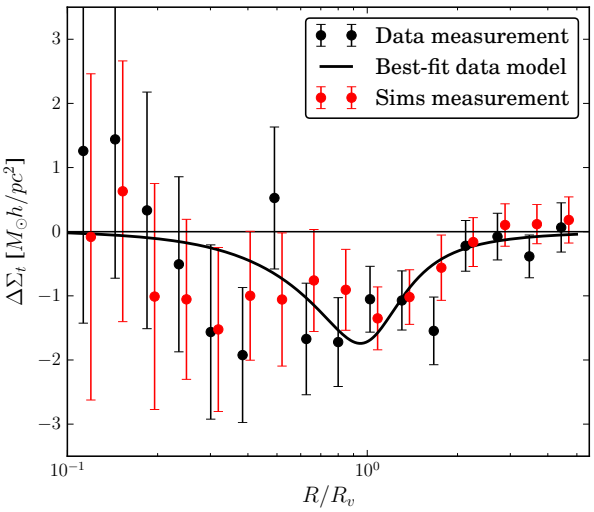

Figure 5.12: Stacked tangential shear profile around voids in DES-SV data (black points) and simulations (red points) in bins of R/R_v. The black solid line shows the best-fit model (see Sect. 5.5.5) to the data shear signal. The χ^2 for the null hypothesis in the data measurement is $\chi^2_{\rm null}/{\rm dof} = 35.5/16$, yielding an estimated $S/N = 4.4$, while the theory model provides a good fit to the data with $\chi^2/{\rm dof}= 13.2/14$. The measurement in the simulations shows consistent with the data best-fit model, yielding $\chi^2/{\rm dof}= 10.1/14$.

The evaluation of this expression yields $\chi^2/\mathrm{dof} = 35.5/16$ and hence $S/N = 4.4$. The significance of the signal is complemented with the null tests in the previous subsection being consistent with the null hypothesis. Furthermore, we test the robustness of the signal to changes in the LOS slicing strategy in Appendix E and to changes in the value of δ_m in Appendix D.

5.5.5 MODEL FITS

We use the 3D void profile of Hamaus, Sutter & Wandelt (2014) (henceforth HSW14)

$$\frac{\rho_\mathrm{v}(r)}{\bar{\rho}} - 1 = \delta_c \frac{1-(r/r_s)^\alpha}{1+(r/R_\mathrm{v})^\beta}\,, \tag{5.13}$$

and fit two parameters: the central underdensity δ_c and the scale radius r_s. Note that r here denotes the 3D (in contrast to projected) radius. We do not fit the inner and outer slopes α and β using the lensing data, but fix their values to the simulation fits of HSW14. That work showed that α and β are not independent parameters but determined by the ratio r_s/R_v, which yields $\alpha = 2.1$ and $\beta = 9.1$ for the best fit r_s shown in Fig. 5.13. Following Krause et al. (2013) the lensing observable $\Delta\Sigma(R/R_v)$ is related to the 3D density by

$$\Delta\Sigma(R/R_v) = \bar{\Sigma}(<R/R_v) - \Sigma(R/R_v)\,, \tag{5.14}$$

where the projected surface density is given by

$$\Sigma(R/R_v) = \int \mathrm{d}r_\mathrm{los}\, \rho_\mathrm{v}\left(\sqrt{r_\mathrm{los}^2 + R^2}\right) - \bar{\rho}\,, \tag{5.15}$$

and $\bar{\rho}$ is the cosmological mean mass density.

The resulting parameter constraints are shown in Fig. 5.13. The reduced $\chi^2/\mathrm{dof} = 13.2/14$ implies a good fit to the theory model. Even though the uncertainties are important, the best-fit $\delta_c = -0.60$ is in agreement with the density profile shown in Fig. 9, which is at the same time in agreement with the profile measured in simulations. In order to further support the data measurement using simulations, we have measured the lensing signal in the simulations using the same number of voids as in the data. The resulting measurement can be found in Fig 5.12, and it shows consistency with the best-fit model to the data with $\chi^2/\mathrm{dof} = 10.1/14$.

Additionally, the best-fit δ_c and the trend in Fig. 5.13 are in agreement with findings in HSW14. However, note the important differences between our work and HSW14: we use photometric galaxies instead of N-body dark matter particles. More importantly, we are using a different void finder. Thus it should not be

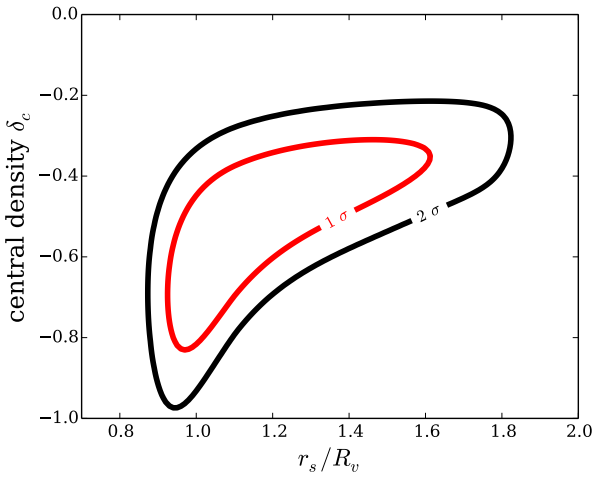

Figure 5.13: Constraints on void central underdensity δ_c and scale radius r_s from the DES-SV data void lensing measurements in Fig. 5.12. Best-fit values are $r_s = 1.05R_v$ and $\delta_c = -0.60$, and the χ^2/dof for the fit is 13.2/14. There is good agreement between the void edge determined from galaxies, R_v, and the void edge determined from lensing, r_s.

surprising that our mean void radius (R_v), scale radius (r_s), and mean void underdensity (δ_c) do not match all the relations obeyed by theirs. For example, their void sample with $r_s/R_v \simeq 1.05$ (matching our best-fit value) is slightly smaller ($R_v \simeq 29$ Mpc/h) and more empty ($\delta_c \simeq -0.7$) than ours.

Finally, we can use the constraints on δ_c being negative as an alternative estimate of the significance in the lensing detection, which is consistent with the estimation in eq. (5.12): marginalizing over r_s, we find $\delta_c < 0$ with a significance of 4.6σ (4.8σ if we fix r_s to its best-fit value). The best-fit value of r_s is compatible with R_v at the 1-σ level. Based on eq. (5.13), $r = r_s$ is just the place where the local 3D density returns to the cosmic mean, $\rho = \bar{\rho}$. The definition of R_v is based on where the local galaxy density returns to the mean (Fig. 1). So given this best-fit model we see that the void wall in the mass distribution (determined from lensing) agrees well with the void wall in the galaxy distribution.

5.5.6 COMPARISON TO PREVIOUS MEASUREMENTS

Other measurements of weak gravitational lensing around voids or underdensities have been performed in recent years. Melchior et al. (2014) used the SDSS void catalog of Sutter et al. (2012) to carry out the first detection of lensing around voids, although at low S/N. Clampitt & Jain (2015), using a similar data sample, optimized the void finding strategy for lensing purposes and were able to achieve a higher $S/N \sim 7$ in the lensing measurement. The void finder in this work is similar to that of Clampitt & Jain (2015), even though we did not attempt to optimise the lensing detection but to minimise the photo-z related impact in the void finding procedure. Our comparable lensing S/N is encouraging given the use of photometric redshifts and a smaller dataset – this highlights the viability of photometric void finders as well as the quality of the DES data.

Gruen et al. (2016) changed the approach and, instead of looking at individual cosmic voids, measured the lensing signal around troughs in the DES-SV galaxy distribution, defined as underdensities in the projection of lens galaxies over a wide range in redshift. That produced a high S/N lensing measurement around those structures, and they succesfully modelled that to probe the connection between galaxies and matter. In that respect, trough lensing does not constrain void profiles or abundances but it is sensitive to the galaxy bias and even cosmology.

5.6 DISCUSSION

We have presented a new void finder designed for photometric surveys and applied it to early Dark Energy Survey data and simulations. Fixing the line-of-sight size of the slice to be at least twice the photo-z scatter, we find the number of voids found in simulated spectroscopic and photometric galaxy catalogs to be within 20% for all transverse void sizes, and indistinguishable for voids with projected size larger than 70 Mpc/h. For such large voids, most have a one-to-one match with nearly the same assigned center and radius.

This result – that the largest voids are the ones most faithfully preserved in a photometric redshift survey – has implications for the expected spatial and dynamic properties of our voids. Ceccarelli et al. (2013) classified voids into those with and without surrounding overdense shells: large voids without shells tend to expand, while smaller voids surrounded by overdense shells are in the process of being crushed by the surrounding shell. This is a useful division for understanding void dynamics, as predicted analytically by Sheth & van de Weygaert (2004) and later studied in simulations (Paz et al., 2013; Ceccarelli et al., 2013; Hamaus, Sutter & Wandelt, 2014) and data (Ruiz et al., 2015). Furthermore, this classification has been useful for predicting large-scale bulk flows of voids in both simulations (Lambas et al., 2016) and data (Ceccarelli et al., 2015). These works found that large voids are on average receding from each other, while small voids in overdense shells are approaching each other.

Most importantly, we have applied the algorithm to the DES-SV data and found a total of 87 voids over the redshift range $0.2 < z < 0.8$. Our $\sim 4\sigma$ detection of the weak gravitational lensing signal of these voids shows they are truly underdense in the matter field and hence not simply a product of Poisson noise, tracer density effects or any systematics in the data. Assuming a model profile (Hamaus, Sutter & Wandelt, 2014), we find a best-fit central density of $\delta_c \sim -0.6$ and scale radius $r_s \sim R_v$. Since r_s is the void edge determined from lensing, and R_v is the edge determined from the galaxy distribution, the best-fit lensing model shows consistency between the mass and galaxy distributions of voids. Note however

that the contours are broad and still allow for the possibility of $r_s \gtrsim R_v$.

Further applications of the same void finder will be explored in future DES data samples. Of particular interest is the study of the CMB cold imprint of voids (Kovács et al. in prep), related to the properties and presence of Dark Energy through the integrated Sachs-Wolfe effect (Granett, Neyrinck & Szapudi, 2008; Cai et al., 2010; Cai, Padilla & Li, 2014; Hotchkiss et al., 2014).

The advances in this work towards finding voids in photometric surveys are also exciting in light of recent advances in void cosmology. Clampitt, Jain & Sánchez (2016) studied void-void and void-galaxy clustering and derived void bias using the spectroscopic SDSS luminous red galaxy (LRG) sample. Hamaus et al. (2016) applied the Alcock-Paczynski test to void clustering statistics to put $\sim 10\%$ constraints on Ω_m using voids identified using CMASS galaxies as tracers, a result that was anticipated in simulations by the same group (Hamaus et al., 2014a,b, 2015). Kitaura et al. (2015) reported greater than 3σ evidence of the presence of baryonic acoustic oscillations (BAO) in void correlations, again using CMASS galaxies. This impressive measurement was made possible by the new void finder presented in Zhao et al. (2015) and detailed studies with mock CMASS samples presented in Liang et al. (2015). While the CMASS sample from BOSS covers a very large area, it lacks a suitable background source sample for direct lensing measurements of void density profiles. Upcoming photometric surveys, which will have many background sources available, will make the combination of void clustering and lensing over large volumes a reality.

In addition to constraining standard cosmological parameters, voids have been used to investigate alternative dark matter scenarios like warm dark matter (Yang et al., 2015), or the effects of neutrinos on void lensing (Massara et al., 2015). Especially numerous are the studies on void abundance (Li, 2011; Clampitt, Cai & Li, 2013; Cai, Padilla & Li, 2015; Zivick et al., 2015; Lam et al., 2015; Pollina et al., 2016) and lensing (Cai, Padilla & Li, 2014; Barreira et al., 2015) as promising probes of alternatives to General Relativity (GR). In particular, Barreira et al. (2015) used simulations of Galileon gravity to show that the lensing signal of voids can be double that in GR. Comparing to the SDSS void lensing results of Clampitt & Jain (2015), they showed that the size of the difference is comparable to current observational errors. Furthermore, another recent development by Cautun, Cai & Frenk (2015) has shown that the signal-to-noise for void lensing can be increased by describing the void profile relative to the boundary rather than the center. Such advances, combined with the increasing quality and volume of data from ongoing surveys, will bring modified gravity constraints using voids within reach. The algorithm in this work ensures that the statistical power of these new photometric datasets can be brought to bear on void measurements.

SUMMARY AND CONCLUSIONS

In this thesis we have studied cosmology from the combination of large-scale structure (LSS) and weak gravitational lensing (WL) using data from the ongoing Dark Energy Survey (DES), producing the first cosmological constraints from the combination of LSS and WL probes in DES, and broadly using only photometric data. For that purpose we have also worked extensively on the characterization of the predominant systematic effects, such as those coming from photometric redshifts and shear estimation mis-calibration. Complementary to using galaxies as lenses, we have also measured the WL signal around cosmic voids, using a void finder developed specifically to be used in photometric surveys, broadening the potential applications of void science as an alternative cosmological probe.

In Chapter 3 we assessed the photometric redshift capabilities of the DES Camera by using an extensive set of photo-z algorithms, and data from the DES Science Verification (SV) period matched to spectroscopic data from surveys like VVDS, zCOSMOS and ACES. That study provides a thorough comparison between training and template-based photo-z methods using real data, and the results for most of the codes surpass the DES requirements on photo-z precision that were defined prior to the start of the survey. In addition, such study sets the basis for subsequent photometric redshift analyses in DES, and the best-performing methods from that comparison are the ones used for the cosmological analysis performed in this thesis, and are still being used nowadays by the DES Collaboration on subsequent data sets.

As the main topic in this thesis, in Chapter 4 we measured the angular clustering and the galaxy-galaxy lensing of red galaxies in the DES-SV data sample and combined them to obtain cosmological constraints. This combination of LSS and WL probes breaks the degeneracies between galaxy bias and cosmological information and it substantially increases the constraining power compared to single probe analyses. In this context, we performed an extensive set of null tests to study potential systematic effects present in the DES-SV data sample. By designing specific tests for each potential systematic, we learned about the impact of the point spread function (PSF) modelling, the survey geometry, mask, and observing

conditions, photometric redshift errors, and possible biases due to source galaxy properties like size or photometric signal-to-noise. The cosmological results of the analysis show consistency with the DES cosmic shear analysis performed using the same underlying data sample. The cosmic shear signal suffers from lower signal-to-noise and different systematics but is sensitive to the matter distribution directly, hence being independent of the complicated galaxy bias.

From the agreement and the complementarity between the analysis in this thesis and the cosmic shear DES analysis, we can conclude that the optimal extraction of cosmological information from photometric galaxy surveys like DES, and future galaxy surveys like LSST, WFIRST and Euclid, will involve the full combination of galaxy clustering, galaxy-galaxy lensing and cosmic shear. This combination not only will maximize the information but also will help in constraining systematic effects that influence each probe differently. Such analyses will have a profound impact in our knowledge about the nature of dark energy, potentially providing hints to new extensions beyond the concordance cosmological model.

Finally, in Chapter 5 we developed a new void finder designed for minimizing the impact of photo-z errors in DES and other photometric surveys, and measured the lensing signal around voids with a significance of more than 4σ using DES-SV data. This measurement is the first detection of void lensing in a photometric survey, and it provides confirmation that the obtained voids are truly underdense in the matter field and hence not a product of Poisson noise, tracer density effects or systematic effects in the data. This void-finder opens the door of void science to the vast amount of data coming from current and future photometric surveys, since while some studies have attempted the extraction of cosmology from cosmic voids, none of them have used data from photometric surveys. Analyses based on cosmic voids are highly complementary to analyses based on the clustering and weak lensing of galaxies. In addition, voids have the advantage of suffering less from non-linearities due to their low-density environment, and therefore they are easier to model and can be of key importance in providing robust and independent cosmological constraints.

Appendix A

DESCRIPTION OF THE METRICS

Here we describe the metrics used throughout this work. For each photo-z code and each galaxy we have either the photo-z estimation and its associated error or a probability density function $P(z)$. As described in the text, a vector of weights were computed in order to match the spectroscopic and photometric samples in multi-color and magnitude space. On each sample we have a vector $\boldsymbol{\omega}$ of weights corresponding to the N galaxies on each test set, where $\sum_{i=1}^{N} \omega_i = 1$. If no weights are used, the default value $\omega_i = \frac{1}{N}$ is assigned to each galaxy. We define the individual bias as $\Delta z_i = z_{\mathrm{phot},i} - z_{\mathrm{spec},i}$ and the statistics used in this work as follows:

1. mean bias($\overline{\Delta z}$):
$$\overline{\Delta z} = \frac{\sum \omega_i \Delta z_i}{\sum \omega_i} \tag{A.1}$$

2. $\sigma_{\Delta z}$:
$$\sigma_{\Delta z} = \left(\frac{\sum \omega_i \left(\Delta z_i - \overline{\Delta z} \right)^2}{\sum \omega_i} \right)^{\frac{1}{2}} \tag{A.2}$$

3. median (Δz_{50}), the median of the Δz distribution, fulfilling:
$$P_{50} = P(\Delta z \leq \Delta z_{50}) = \int_0^{\Delta z_{50}} \omega(\Delta z) d(\Delta z) = \frac{1}{2} \tag{A.3}$$

4. σ_{68}, half of the width of the distribution, measured with respect to the median, where 68% of the data are enclosed. This is computed as:
$$\sigma_{68} = \frac{1}{2} \left(P_{84} - P_{16} \right) \tag{A.4}$$

5. $\text{out}_{2\sigma}$, the fraction of outliers above the $2\sigma_{\Delta z}$ level:

$$\text{out}_{2\sigma} = \frac{\sum W_i}{\sum \omega_i} \tag{A.5}$$

where,

$$W_i = \begin{cases} \omega_i, & \text{if } |\Delta z_i - \overline{\Delta z}| > 2\sigma_{\Delta z} \\ 0, & \text{if } |\Delta z_i - \overline{\Delta z}| \leq 2\sigma_{\Delta z} \end{cases}$$

6. $\text{out}_{3\sigma}$, the fraction of outliers above the $3\sigma_{\Delta z}$ level:

$$\text{out}_{3\sigma} = \frac{\sum W_i}{\sum \omega_i} \tag{A.6}$$

where,

$$W_i = \begin{cases} \omega_i, & \text{if } |\Delta z_i - \overline{\Delta z}| > 3\sigma_{\Delta z} \\ 0, & \text{if } |\Delta z_i - \overline{\Delta z}| \leq 3\sigma_{\Delta z} \end{cases}$$

7. $\overline{\Delta z'}$, the mean of the distribution of Δz is normalized by their estimated errors. Ideally this distribution should resemble a normal distribution with zero mean and unit variance. We define $\Delta z'_i = \Delta z_i / \epsilon_{\text{phot},i}$ where $\epsilon_{\text{phot},i}$ is the computed error of the photometric redshift for galaxy i. Then:

$$\overline{\Delta z'} = \frac{\sum \omega_i \Delta z'_i}{\sum \omega_i} \tag{A.7}$$

8. $\sigma_{\Delta z'}$:

$$\sigma_{\Delta z'} = \left(\frac{\sum \omega_i \left(\Delta z'_i - \overline{\Delta z'} \right)^2}{\sum \omega_i} \right)^{\frac{1}{2}} \tag{A.8}$$

9. N_{poisson}, a metric that quantifies how close the distribution of photometric redshifts $N(z_{\text{phot}})$ is to the distribution of spectroscopic redshifts $N(z_{\text{spec}})$. For each photometric redshift bin j of width 0.1, we compute the difference of $N(z_{\text{phot}}) - N(z_{\text{spec}})$ normalized by the Poisson fluctuations on $N(z_{\text{spec}})$:

$$n_{\text{poisson},j} = \frac{\left(\sum_{z_{\text{phot},i} \,\epsilon\, \text{bin}_j} \omega_i N - \sum_{z_{\text{spec},i} \,\epsilon\, \text{bin}_j} \omega_i N \right)}{\sqrt{\sum_{z_{\text{spec},i} \,\epsilon\, \text{bin}_j} \omega_i N}}$$

Then N_{poisson} is computed as the RMS of the previous quantity:

$$N_{\text{poisson}} = \left(\frac{1}{n_{bins}} \sum_{j=1}^{n_{bins}} n^2_{\text{poisson},j} \right)^{\frac{1}{2}} \tag{A.9}$$

10. KS is the Kolmogorov-Smirnov test that quantifies whether the two redshift distributions ($N(z_{\text{phot}})$ and $N(z_{\text{spec}})$) are compatible with being drawn from the same parent distribution, independently of binning. It is defined as the maximum distance between both empirical cumulative distributions. The lower this value, the closer are both distributions. The empirical cumulative distribution function is calculated as:

$$F_{\text{spec}}(z) = \frac{\sum_{i=1}^{N} \Omega_{z_{\text{spec},i}<z}}{\sum \omega_i}$$

where,

$$\Omega_{z_{\text{spec},i}<z} = \begin{cases} \omega_i, & \text{if } z_{\text{spec},i} < z \\ 0, & \text{otherwise} \end{cases}$$

Similarly, the empirical cumulative distribution function $F_{\text{phot}}(z)$ is computed for $N(z_{\text{phot}})$. Then the KS statistic is computed as:

$$\text{KS} = \max_z \left(|F_{\text{phot}}(z) - F_{\text{spec}}(z)| \right) \tag{A.10}$$

For the submissions with a $P(z)$ for each galaxy, these cumulative distributions are computed taking into account the $p(z)$ distribution for each galaxy.

Table A.1: Results of all the photo-z metrics listed in Appendix A for all the codes analyzed in Test 1. The errors are computed from bootstrap resampling with 100 samples. The weighting procedure has been applied, together with a cut on the 10% of the galaxies with the highest estimated photo-z errors for each code.

Test 1	$\overline{\Delta z}$	Δz_{50}	σ_{68}	$\sigma_{\Delta z}$	$\text{out}_{2\sigma}$	$\text{out}_{3\sigma}$	$\overline{\Delta z\prime}$	$\sigma_{\Delta z\prime}$	$\text{N}_{\text{poisson}}$	KS
DESDM	-0.005 ± 0.003	-0.003 ± 0.002	0.094 ± 0.002	0.135 ± 0.005	0.053 ± 0.005	0.018 ± 0.003	-0.047 ± 0.032	1.47 ± 0.05	7.03 ± 0.48	0.056 ± 0.004
ANNz	0.002 ± 0.003	-0.001 ± 0.003	0.086 ± 0.002	0.118 ± 0.004	0.049 ± 0.004	0.015 ± 0.002	0.096 ± 0.046	3.34 ± 0.13	6.35 ± 0.48	0.052 ± 0.005
TPZ	-0.001 ± 0.003	0.004 ± 0.002	0.078 ± 0.002	0.122 ± 0.006	0.046 ± 0.004	0.019 ± 0.002	0.019 ± 0.032	1.52 ± 0.06	4.12 ± 0.50	0.044 ± 0.005
RVMz	-0.011 ± 0.005	-0.004 ± 0.002	0.116 ± 0.004	0.180 ± 0.008	0.060 ± 0.005	0.023 ± 0.003	-0.084 ± 0.041	1.37 ± 0.09	8.38 ± 0.54	0.083 ± 0.007
NIP-kNNz	-0.030 ± 0.005	-0.011 ± 0.002	0.120 ± 0.004	0.197 ± 0.009	0.058 ± 0.005	0.018 ± 0.003	-0.186 ± 0.030	1.11 ± 0.11	3.63 ± 0.48	0.054 ± 0.007
ANNz2	-0.002 ± 0.003	-0.003 ± 0.002	0.089 ± 0.003	0.151 ± 0.009	0.042 ± 0.003	0.021 ± 0.002	0.063 ± 0.071	2.28 ± 0.25	5.38 ± 0.45	0.052 ± 0.004
BPZ	-0.022 ± 0.003	-0.021 ± 0.002	0.097 ± 0.003	0.137 ± 0.006	0.049 ± 0.003	0.018 ± 0.002	-0.194 ± 0.032	1.75 ± 0.07	7.91 ± 0.62	0.099 ± 0.008
EAZY	-0.061 ± 0.004	-0.063 ± 0.003	0.109 ± 0.003	0.153 ± 0.010	0.035 ± 0.005	0.015 ± 0.002	-0.331 ± 0.074	3.98 ± 0.80	11.14 ± 0.86	0.195 ± 0.009
LePhare	0.002 ± 0.004	-0.007 ± 0.003	0.111 ± 0.003	0.171 ± 0.008	0.047 ± 0.002	0.024 ± 0.002	1.177 ± 0.379	42.88 ± 6.60	6.63 ± 0.44	0.087 ± 0.006
PhotoZ	-0.029 ± 0.003	-0.029 ± 0.002	0.097 ± 0.003	0.142 ± 0.006	0.058 ± 0.004	0.017 ± 0.002	-0.268 ± 0.020	1.00 ± 0.03	4.56 ± 0.464	0.080 ± 0.006
TPZ P(z)	0.006 ± 0.003	0.011 ± 0.002	0.078 ± 0.002	0.119 ± 0.006	0.049 ± 0.005	0.018 ± 0.003	0.125 ± 0.030	1.48 ± 0.05	2.60 ± 0.44	0.044 ± 0.006
ArborZ P(z)	-0.008 ± 0.004	0.001 ± 0.004	0.128 ± 0.003	0.153 ± 0.005	0.056 ± 0.005	0.016 ± 0.003	-0.028 ± 0.024	0.96 ± 0.02	4.77 ± 0.44	0.064 ± 0.009
ANNz2 P(z)	-0.002 ± 0.002	-0.002 ± 0.002	0.085 ± 0.002	0.118 ± 0.004	0.051 ± 0.004	0.016 ± 0.003	-0.004 ± 0.024	1.31 ± 0.04	5.01 ± 0.40	0.045 ± 0.004
SkyNet P(z)	-0.001 ± 0.002	0.001 ± 0.002	0.077 ± 0.002	0.104 ± 0.003	0.072 ± 0.006	0.015 ± 0.002	-0.006 ± 0.014	0.82 ± 0.02	4.09 ± 0.40	0.052 ± 0.005
BPZ P(z)	-0.025 ± 0.003	-0.025 ± 0.003	0.101 ± 0.002	0.132 ± 0.006	0.046 ± 0.004	0.014 ± 0.002	-0.224 ± 0.033	1.75 ± 0.080	5.77 ± 0.94	0.098 ± 0.007
ZEBRA P(z)	-0.050 ± 0.005	-0.030 ± 0.002	0.109 ± 0.004	0.177 ± 0.012	0.043 ± 0.006	0.018 ± 0.002	-0.383 ± 0.047	1.90 ± 0.16	6.69 ± 0.55	0.112 ± 0.008

Table A.2: Results of all the photo-z metrics for all the codes analyzed in Test 2. The errors are computed from bootstrap resampling with 100 samples. The weighting procedure has been applied, together with a cut on the 10% of the galaxies with the highest estimated photo-z errors for each code.

Test 2	$\overline{\Delta z}$	Δz_{50}	σ_{68}	$\sigma_{\Delta z}$	$\text{out}_{2\sigma}$	$\text{out}_{3\sigma}$	$\overline{\Delta z\prime}$	$\sigma_{\Delta z\prime}$	N_{poisson}	KS
DESDM	0.001 ± 0.003	-0.002 ± 0.002	0.087 ± 0.003	0.123 ± 0.005	0.052 ± 0.006	0.016 ± 0.003	0.008 ± 0.041	1.64 ± 0.05	8.82 ± 0.71	0.058 ± 0.008
ANNz	-0.001 ± 0.003	-0.003 ± 0.002	0.075 ± 0.002	0.105 ± 0.007	0.043 ± 0.006	0.013 ± 0.002	0.013 ± 0.062	3.22 ± 0.09	7.69 ± 0.67	0.057 ± 0.006
TPZ	0.005 ± 0.003	0.004 ± 0.002	0.066 ± 0.002	0.101 ± 0.005	0.042 ± 0.004	0.022 ± 0.003	0.091 ± 0.042	1.63 ± 0.09	4.12 ± 0.48	0.045 ± 0.007
RVMz	-0.008 ± 0.005	-0.005 ± 0.003	0.110 ± 0.004	0.168 ± 0.008	0.063 ± 0.006	0.016 ± 0.002	-0.043 ± 0.041	1.28 ± 0.14	10.16 ± 0.75	0.079 ± 0.008
NIP-kNNz	-0.030 ± 0.005	-0.013 ± 0.003	0.107 ± 0.005	0.169 ± 0.009	0.060 ± 0.005	0.021 ± 0.003	-0.291 ± 0.076	2.04 ± 0.42	4.46 ± 0.65	0.054 ± 0.007
ANNz2	0.001 ± 0.005	-0.005 ± 0.002	0.083 ± 0.003	0.140 ± 0.010	0.048 ± 0.007	0.017 ± 0.002	0.094 ± 0.080	2.78 ± 0.45	7.73 ± 0.70	0.051 ± 0.006
BPZ	-0.009 ± 0.004	-0.015 ± 0.002	0.096 ± 0.003	0.143 ± 0.007	0.054 ± 0.005	0.017 ± 0.002	0.044 ± 0.062	2.40 ± 0.16	8.20 ± 0.75	0.053 ± 0.005
EAZY	-0.071 ± 0.004	-0.069 ± 0.003	0.094 ± 0.003	0.155 ± 0.010	0.048 ± 0.005	0.020 ± 0.003	-0.650 ± 0.171	12.35 ± 3.23	11.07 ± 0.83	0.210 ± 0.010
LePhare	0.019 ± 0.005	-0.008 ± 0.002	0.099 ± 0.003	0.183 ± 0.010	0.053 ± 0.004	0.028 ± 0.004	2.099 ± 0.994	120.10 ± 19.65	7.50 ± 1.16	0.071 ± 0.007
PhotoZ	-0.019 ± 0.004	-0.029 ± 0.003	0.093 ± 0.003	0.147 ± 0.007	0.053 ± 0.005	0.022 ± 0.004	1.514 ± 0.829	39.65 ± 10.27	4.93 ± 0.60	0.054 ± 0.007
TPZ P(z)	0.011 ± 0.002	0.011 ± 0.002	0.067 ± 0.003	0.099 ± 0.004	0.042 ± 0.005	0.021 ± 0.003	0.210 ± 0.038	1.57 ± 0.09	3.70 ± 0.45	0.051 ± 0.006
ArborZ P(z)	-0.003 ± 0.003	0.001 ± 0.002	0.096 ± 0.003	0.119 ± 0.006	0.055 ± 0.006	0.010 ± 0.002	-0.017 ± 0.023	0.90 ± 0.04	5.47 ± 0.49	0.055 ± 0.004
ANNz2 P(z)	-0.003 ± 0.003	-0.003 ± 0.002	0.076 ± 0.002	0.110 ± 0.007	0.053 ± 0.008	0.013 ± 0.004	-0.032 ± 0.041	1.57 ± 0.08	6.68 ± 0.46	0.050 ± 0.005
SkyNet P(z)	0.004 ± 0.003	-0.002 ± 0.002	0.064 ± 0.002	0.110 ± 0.008	0.047 ± 0.006	0.016 ± 0.004	0.017 ± 0.035	1.20 ± 0.09	9.80 ± 0.70	0.079 ± 0.006
BPZ P(z)	-0.009 ± 0.005	-0.019 ± 0.002	0.097 ± 0.003	0.146 ± 0.008	0.048 ± 0.004	0.019 ± 0.003	-0.020 ± 0.067	2.46 ± 0.23	6.16 ± 0.88	0.058 ± 0.007
ZEBRA P(z)	-0.027 ± 0.005	-0.019 ± 0.003	0.103 ± 0.003	0.163 ± 0.008	0.054 ± 0.005	0.016 ± 0.003	-0.292 ± 0.075	4.24 ± 0.64	8.40 ± 1.42	0.071 ± 0.007

Table A.3: Results of all the photo-z metrics for all the codes analyzed in Test 3. The errors are computed from bootstrap resampling with 100 samples. The weighting procedure has been applied, together with a cut on the 10% of the galaxies with the largest estimated photo-z errors for each code.

Test 3	$\overline{\Delta z}$	Δz_{50}	σ_{68}	$\sigma_{\Delta z}$	$\text{out}_{2\sigma}$	$\text{out}_{3\sigma}$	$\overline{\Delta z\prime}$	$\sigma_{\Delta z\prime}$	N_{poisson}	KS
DESDM	-0.010 ± 0.003	-0.010 ± 0.002	0.091 ± 0.002	0.127 ± 0.004	0.052 ± 0.005	0.015 ± 0.002	-0.111 ± 0.027	1.42 ± 0.04	6.81 ± 0.52	0.076 ± 0.006
ANNz	-0.006 ± 0.002	-0.009 ± 0.002	0.085 ± 0.002	0.114 ± 0.003	0.044 ± 0.004	0.017 ± 0.003	-0.063 ± 0.036	2.66 ± 0.07	7.86 ± 0.55	0.096 ± 0.007
TPZ	-0.004 ± 0.003	-0.001 ± 0.002	0.076 ± 0.002	0.124 ± 0.006	0.048 ± 0.004	0.021 ± 0.003	-0.047 ± 0.040	1.77 ± 0.06	3.81 ± 0.83	0.038 ± 0.005
RVMz	-0.033 ± 0.005	-0.016 ± 0.003	0.118 ± 0.005	0.177 ± 0.009	0.050 ± 0.004	0.019 ± 0.003	-0.186 ± 0.029	1.04 ± 0.05	8.10 ± 0.59	0.088 ± 0.008
NIP-kNNz	-0.039 ± 0.005	-0.011 ± 0.003	0.115 ± 0.004	0.191 ± 0.009	0.053 ± 0.004	0.020 ± 0.002	-0.270 ± 0.047	1.88 ± 0.51	3.92 ± 0.37	0.068 ± 0.007
ANNz2	-0.014 ± 0.004	-0.012 ± 0.002	0.101 ± 0.004	0.144 ± 0.005	0.056 ± 0.004	0.019 ± 0.003	-0.087 ± 0.063	2.66 ± 0.20	6.51 ± 0.51	0.071 ± 0.007
BPZ	-0.021 ± 0.003	-0.021 ± 0.002	0.098 ± 0.003	0.138 ± 0.006	0.049 ± 0.004	0.018 ± 0.002	-0.182 ± 0.032	1.75 ± 0.07	7.89 ± 0.62	0.096 ± 0.008
LePhare	0.004 ± 0.004	-0.008 ± 0.003	0.111 ± 0.003	0.170 ± 0.007	0.047 ± 0.003	0.024 ± 0.002	1.599 ± 0.429	50.48 ± 6.96	6.37 ± 0.43	0.089 ± 0.006
PhotoZ	-0.031 ± 0.003	-0.032 ± 0.002	0.097 ± 0.003	0.140 ± 0.006	0.057 ± 0.004	0.017 ± 0.002	-0.294 ± 0.020	1.00 ± 0.02	4.99 ± 0.51	0.088 ± 0.007
TPZ P(z)	0.002 ± 0.003	0.006 ± 0.002	0.076 ± 0.002	0.119 ± 0.006	0.049 ± 0.005	0.020 ± 0.003	0.082 ± 0.038	1.74 ± 0.07	3.32 ± 0.36	0.030 ± 0.004
ArborZ P(z)	-0.016 ± 0.003	-0.011 ± 0.003	0.105 ± 0.002	0.128 ± 0.005	0.053 ± 0.004	0.011 ± 0.002	-0.112 ± 0.023	0.95 ± 0.03	4.95 ± 0.38	0.081 ± 0.006
ANNz2 P(z)	-0.025 ± 0.004	-0.015 ± 0.002	0.088 ± 0.003	0.128 ± 0.007	0.049 ± 0.004	0.018 ± 0.003	-4.557 ± 0.712	26.89 ± 2.97	7.40 ± 0.54	0.098 ± 0.008
SkyNet P(z)	0.007 ± 0.004	0.007 ± 0.003	0.088 ± 0.002	0.129 ± 0.006	0.068 ± 0.005	0.023 ± 0.003	0.071 ± 0.025	0.98 ± 0.03	8.14 ± 0.86	0.093 ± 0.006
BPZ P(z)	-0.024 ± 0.003	-0.024 ± 0.003	0.101 ± 0.002	0.130 ± 0.006	0.046 ± 0.004	0.014 ± 0.002	-0.211 ± 0.033	1.73 ± 0.07	5.32 ± 0.50	0.096 ± 0.007
ZEBRA P(z)	-0.050 ± 0.005	-0.030 ± 0.002	0.109 ± 0.004	0.177 ± 0.012	0.043 ± 0.006	0.018 ± 0.002	-0.383 ± 0.047	1.90 ± 0.16	6.69 ± 0.55	0.112 ± 0.008

Appendix B

Excess Surface Density $\Delta\Sigma$

In this section, we present complementary cosmology results obtained for the fiducial redMaGiC lens bin ($0.35 < z < 0.50$) by using the excess surface density, $\Delta\Sigma(R)$ as a proxy for the galaxy-galaxy lensing signal of redMaGiC galaxies. For this purpose, we define another lensing estimator that optimally weights each lens-source pair of galaxies depending on the line-of-sight distance separating them. This effectively downweights pairs of galaxies which are very close and for which we expect a small lensing efficiency. The observable is estimated from the measured shapes of background galaxies as

$$\Delta\Sigma^{\rm lens}(R; z_{\rm L}) = \frac{\sum_j \left[\omega'_j \gamma_{t,j}(R)/\Sigma^{-1}_{{\rm crit},j}(z_{\rm L}, z_{\rm s})\right]}{\sum_j \omega'_j} \tag{B.1}$$

where the summation $\sum_j$ goes over all the source galaxies in the radial bin R, around all the lens galaxy positions, and the weight factor for the j-th galaxy is given by

$$\omega'_j = \omega_j \, \Sigma^{-2}_{{\rm crit},j}(z_{\rm L}, z_{\rm s}) \,. \tag{B.2}$$

Note that, in contrast with $\gamma_t(\theta)$, for $\Delta\Sigma$ we bin source galaxies according to radial distance R in the region around each lens galaxy, instead of angular scale θ. In order to estimate distances, we assume a flat ΛCDM model with $\Omega_m = 0.3$. The weighting factor $\Sigma_{\rm crit}(z_{\rm L}, z_{\rm s})$ is computed as a function of lens and source redshifts for the assumed cosmology as

$$\Sigma_{\rm crit}(z_{\rm L}, z_{\rm s}) = \frac{c^2}{4\pi G} \frac{D_A(z_{\rm s})}{D_A(z_{\rm L}) D_A(z_{\rm L}, z_{\rm s})} \,, \tag{B.3}$$

where $\Sigma^{-1}_{\rm crit}(z_{\rm L}, z_{\rm s}) = 0$ for $z_{\rm s} < z_{\rm L}$ and D_A is the angular diameter distance. We have checked that changes in the assumed cosmology have little impact in the es-

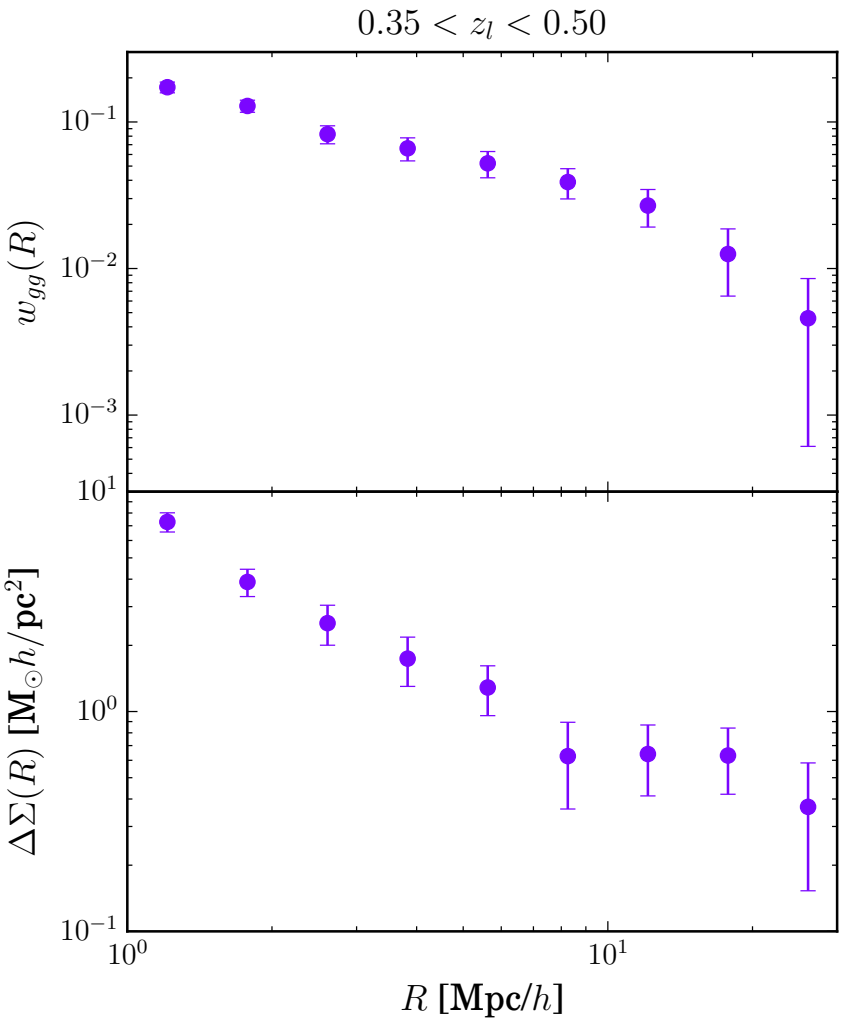

Figure B.1: Same as our fiducial measurement plot in Fig. 4.2, but using the alternative lensing estimator $\Delta\Sigma$. In addition, the data are binned with respect to projected physical distance (R [Mpc/h]) rather than angle (θ [arcmins]). The measurements are very similar to our fiducial results, as are the corresponding cosmological constraints in Fig. 4.6.

timation of $\Delta\Sigma$ so that they are not relevant for the analysis presented in this work (see also Mandelbaum et al. 2013). Finally, just as we do with tangential shear measurements, our final estimator involves subtracting the contribution around random points, to which now we assign redshifts randomly drawn from the real lens redshift distribution.

Figure B.1 shows the clustering and the galaxy-galaxy lensing signals, the latter using the alternative $\Delta\Sigma$ estimator, both binned according to projected radial distance R around lenses. In this case, we use all source galaxies available in the `ngmix` fiducial shear catalog and we weight each lens-source galaxy pair according to their individual photometric redshifts so that nearby pairs for which we expect a small lensing efficiency are effectively downweighted. For the angular clustering, essentially the same dataset is used in Fig. B.1 as for our fiducial results pictured in Fig. 4.2. Thus, the two plots are very similar, with the main difference being the range of scales shown on the x-axis.

Our cosmological constraints obtained from fitting for $\Delta\Sigma(R)$ and $w(R)$ are shown in Fig. 4.6. These are consistent with our fiducial results, and show tighter constraints on parameters like Ω_m, due to the optimal lens weighting and the larger number of source galaxies effectively used. However, we do not use this estimator as the fiducial since we follow the photo-z error modelling of Bonnett et al. (2015) and The Dark Energy Survey Collaboration (2015), where the nuisance parameters act as an overall shift in the full stacked distribution instead of on a galaxy-by-galaxy basis. Our choice of estimator does not seem to have a large impact on the constraints derived from our analysis (see Fig. 4.6).

Appendix C

ngmix vs. im3shape

In Section 4.5.3 we studied the consistency of the obtained cosmological constraints when using the two shear pipelines presented in Jarvis et al. (2015). In Fig. C.1 we show the actual comparison of the lensing measurements from the two shear pipelines, for all the different lens - source bin configurations. The `im3shape` results are an excellent match to our fiducial measurements with `ngmix` (shown earlier in Fig. 4.2).

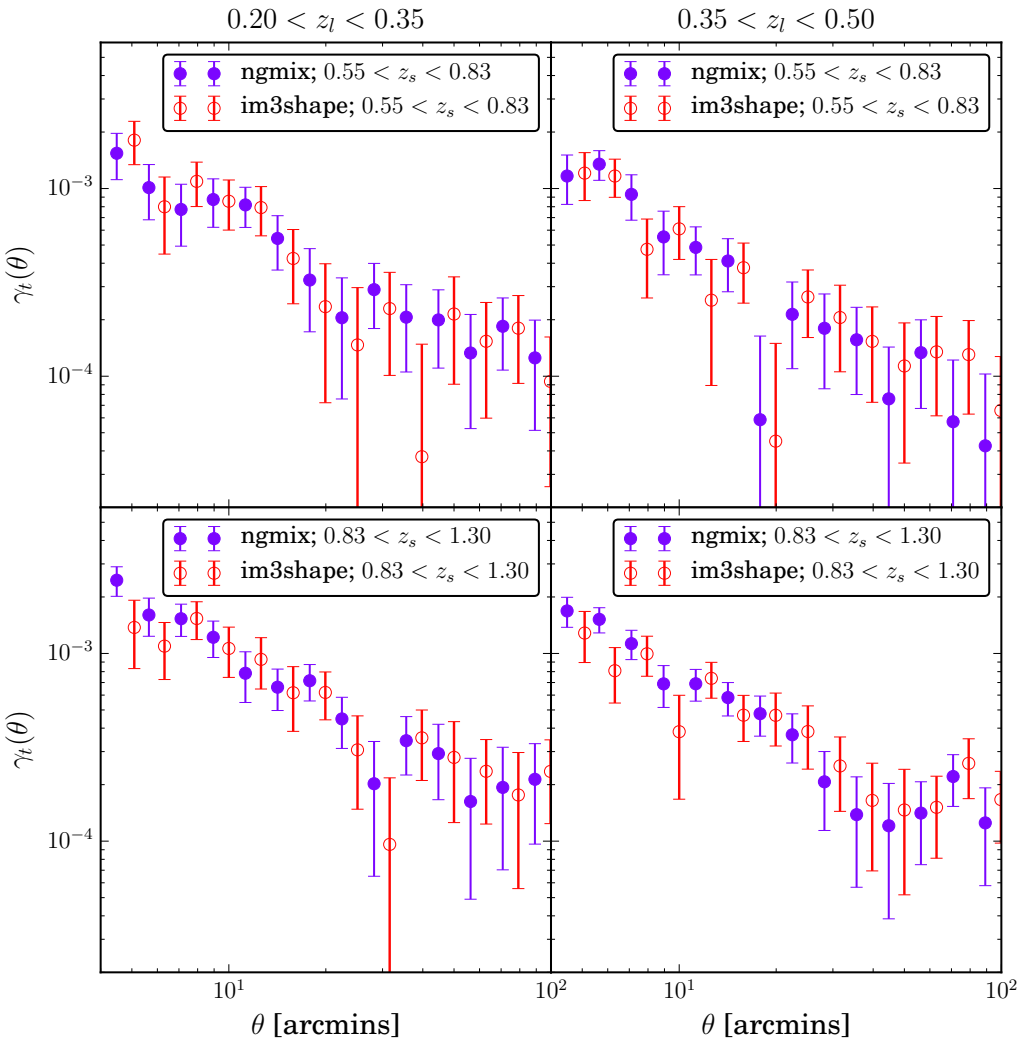

Figure C.1: Comparison of the tangential shear signal using `ngmix` (solid purple circles) and `im3shape` (open red circles) shear pipelines. The result is shown for the two lens redshift bins (left and right columns) and the two source redshift bins (upper and lower rows) used in this work. For all bin combinations, the agreement between pipelines is excellent.

Appendix D

Choice of δ_m

The void finder presented in Sect. 5.3 of this chapter produces a void catalog which depends on the chosen value for the maximum density contrast (δ_m) of a pixel to become a void center (see Sect. 5.3.1). The most significant, and hence deepest voids found by the algorithm are independent of the choice of δ_m, but the total number of voids in the catalog will vary with that choice. With the fiducial value being $\delta_m = -0.30$, in this appendix we vary that value by 10% high and low, and test the impact of these changes in the void lensing signal in the data.

The fiducial void catalog with $\delta_m = -0.30$ contains 78 voids and the goodness of the best-fit model to its lensing signal (see Sect. 5.5.5) is 13.2/14. The catalog with $\delta_m = -0.33$ contains 73 voids and the goodness of the lensing fiducial best-fit model is 12.9/14. The catalog with $\delta_m = -0.27$ contains 107 voids and the goodness of the lensing fiducial best-fit model is 11.9/14. The good agreement between the lensing signal in the three cases is also shown in Fig. D.1.

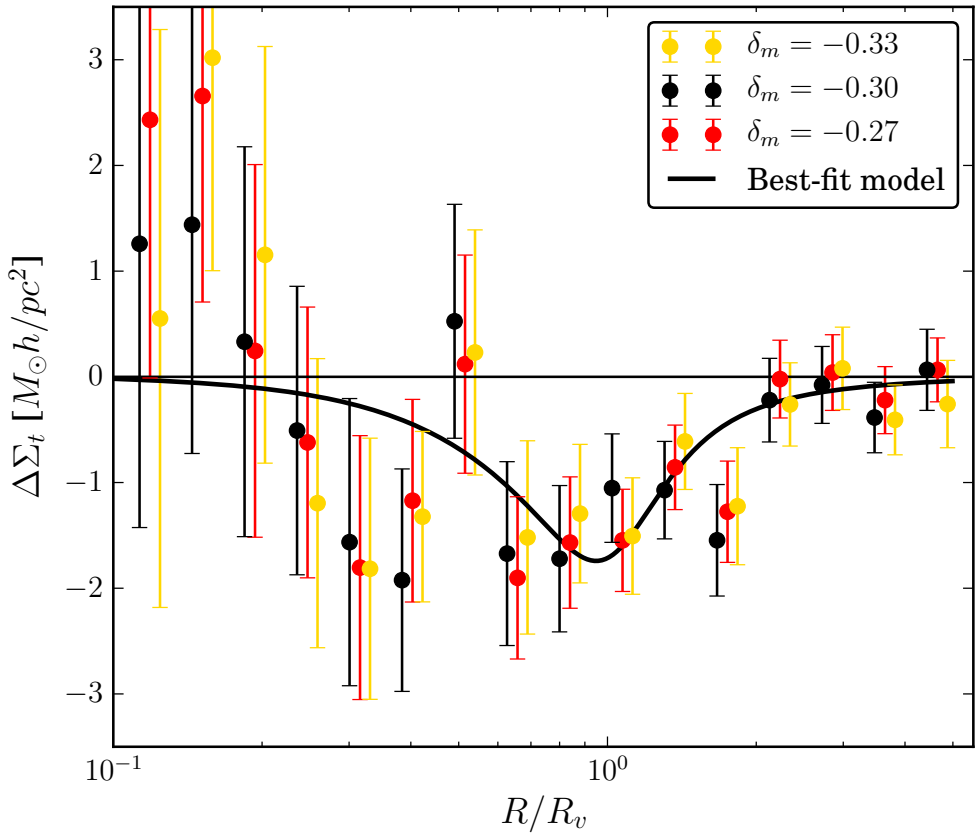

Figure D.1: Stacked void lensing signal in DES-SV data for three choices of δ_m: −0.33, −0.30 (fiducial), −0.27. The black line shows the best-fit model to the fiducial measurement. The comparison shows good agreement between the three sets of measurements.

Appendix E

LENSING ON INDIVIDUAL SLICINGS

In Sect. 5.4.2 we presented a way of combining different slicings of the line of sight (LOS), oversampling it with slices of 100 Mpc/h thickness every 20 Mpc/h, in order to get more information in that direction. Voids found in neighboring slices are joined if their centers are close enough, and the resulting group of voids is considered an individual physical underdensity.

In this appendix we test the impact of that procedure on the void lensing results presented in this chapter (Sect. 5.5). For that purpose, we perform the lensing measurement on the set of voids found in each individual slicing, corresponding to the five columns in the graphical representation of Fig. 5.6. Note that in the case of individual slicings there is no overlap between the slices in which voids are found. The corresponding five lensing measurements, together with its mean and standard deviation, are shown in Fig. E.1, where they are compared to the lensing measurement presented in Sect. 5.5. The comparison in that plot, with the majority of points from the combined slicings measurement being within 1σ of the mean individual slicings case, shows how the combined slicing approach is not affecting the lensing results in this work in any other way than reducing the noise in the measurement.

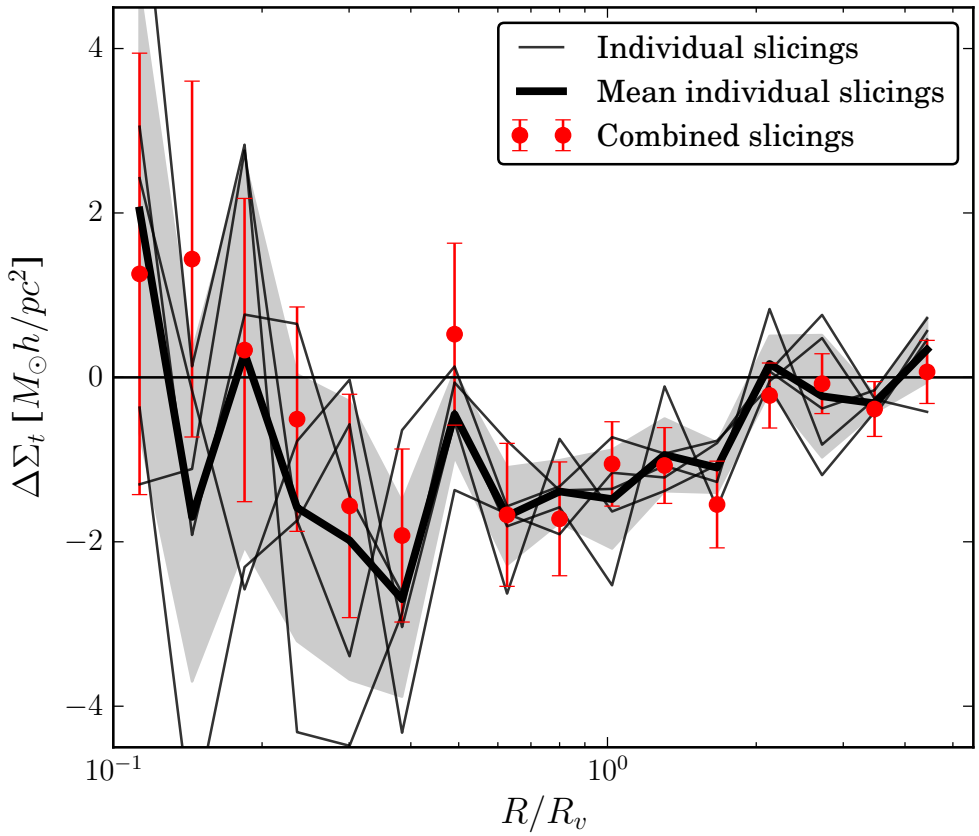

Figure E.1: Stacked void lensing signal in DES-SV data for each of the five individual slicings (thin black lines) and for their mean (thick black line), compared to the standard deviation of the individual slicings measurements (shaded grey region). The actual measurement of the final void catalog from Sect. 5.5 is also shown (red data points with errors). This comparison shows good agreement between the combined and individual slicings.

Appendix F

Randomized VOID CATALOG

The *randomized* void catalog in this chapter is produced such that it mimics the properties of the true void catalog in redshift and radius. We start from a set of random points inside the data mask; they will constitute the centers of the *randomized* voids. We assign a redshift to each random point drawn for the true redshift distribution of voids and, to each *randomized* void, we assign an angular radius from the true distribution of angular radii for voids of similar redshift (in a window of $\Delta z = 0.1$), this way preserving the redshift - angular radius relation. Finally, from the angular radius and the redshift we compute the comoving radius of the *randomized* voids.

After this process we have a *randomized* void catalog with the same properties as the true one. Then, we also apply the process described in Sect. 5.4.1 to get rid of voids near the survey edges. At the end, the *randomized* void catalog has 10 times as many objects as the true one. Figure F.1 shows the agreement between the distributions of the true and *randomized* voids in redshift and comoving and angular radius.

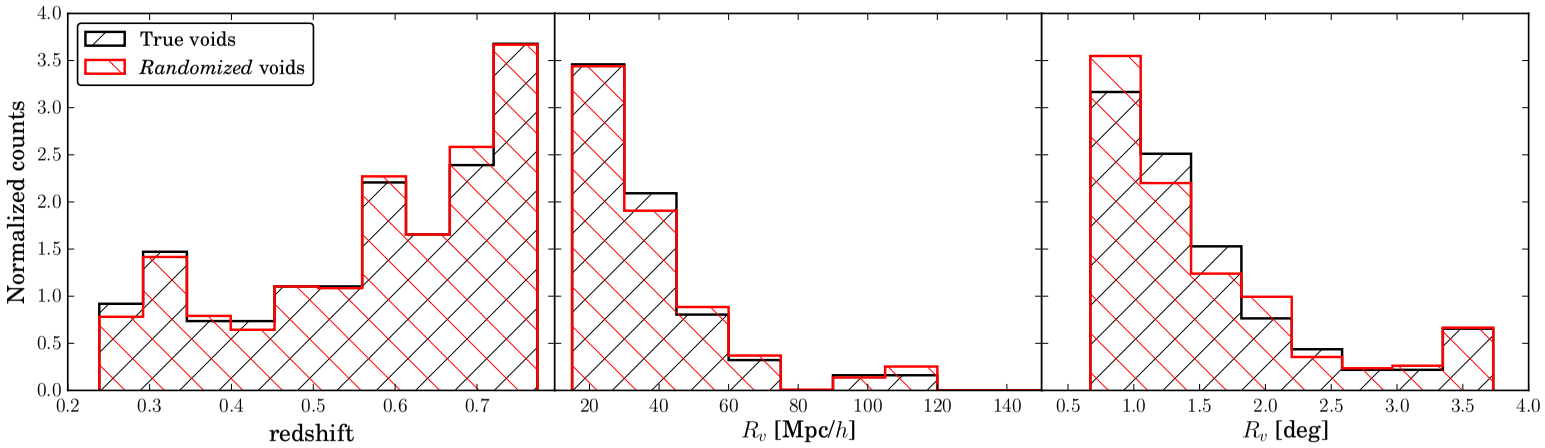

Figure F.1: Comparison of the true and *randomized* void redshift (*left panel*), comoving radius (*center panel*) and angular radius distributions (*right panel*). The *randomized* void catalog is produced to mimic these properties of the true void catalog by following the procedure explained in Appendix F.

BIBLIOGRAPHY

Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, Mon. Not. R. Astron. Soc., 417, 1891

Ahn C. P. et al., 2012, Astrophys. J. Suppl. Ser., 203, 21

Aihara H. et al., 2011, Astrophys. J. Suppl. Ser., 193, 29

Anderson L. et al., 2014, Mon. Not. R. Astron. Soc., 441, 24

Arnouts S. et al., 2002, Mon. Not. R. Astron. Soc., 329, 355

Banerji M., Abdalla F. B., Lahav O., Lin H., 2008, Mon. Not. R. Astron. Soc., 386, 1219

Barreira A., Cautun M., Li B., Baugh C. M., Pascoli S., 2015, J. Cosmol. Astropart. Phys., 2015, 028

Becker M. R. et al., 2015, arXiv:1507.05598

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, Astrophys. J., 762, 109

Bender R., Appenzeller I., Bohm A., al E., 2001, Deep Fields, ed. S. Cris. A. Renzini, R. E. Williams,, 96

Benitez N., 2000, Astrophys. J., 536, 571

Bertin E., 2006, ASP Conf. Ser. 351, Astron. Data Anal. Softw. Syst. XV, ed. C. Gabriel, C. Arviset, D. Ponz, E. Solano (San Fr. CA ASP), 112

Bertin E., 2011, Publ. Astron. Data Anal. Softw. Syst. XX. ASP Conf. Proceedings, Vol. 442, Proc. a Conf. held Seapt. World Trade Center, Boston, Massachusetts, USA 7-11 Novemb. 2010. Ed. by Ian N. Evans, Alberto Accoma, 435

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, ASP Conf. Proc. 281, Astron. Data Anal. Softw. Syst. XI, ed. D. A. Bohlender, D. Durand, T. H. Handley (San Fr. CA ASP), 228

Betancort-Rijo J., Patiri S. G., Prada F., Romano A. E., 2009, Mon. Not. R. Astron. Soc., 400, 1835

Bolzonella M., Miralles J., Pell R., 2000, Astron. Astrophys., 492, 476

Bonnett C., 2013, ArXiv e-prints

Bonnett C. et al., 2015, arXiv:1507.05909

Bordoloi R., Lilly S. J., Amara A., 2010, Mon. Not. R. Astron. Soc., no

Bos E. G. P., van de Weygaert R., Dolag K., Pettorino V., 2012, Mon. Not. R. Astron. Soc., 426, 440

Brammer G. B., van Dokkum P. G., Coppi P., 2008, Astrophys. J., 686, 1503

Brimioulle F., Seitz S., Lerchster M., Bender R., Snigula J., 2013, Mon. Not. R. Astron. Soc., 432, 1046

Bruzual G., Charlot S., 2003, Mon. Not. R. Astron. Soc., 344, 1000

Bruzual A. G., Charlot S., 1993, Astrophys. J., 405, 538

Busha M. T., Wechsler R. H., Becker M. R., Erickson B., Evrard A. E., 2013, Am. Astron. Soc. AAS Meet. #221, id.341.07

Cai Y.-C., Cole S., Jenkins A., Frenk C. S., 2010, Mon. Not. R. Astron. Soc., 407, 201

Cai Y.-C., Padilla N., Li B., 2014, arXiv:1410.8355

Cai Y.-C., Padilla N., Li B., 2015, Mon. Not. R. Astron. Soc., 451, 1036

Carrasco Kind M., Brunner R. J., 2013, Mon. Not. R. Astron. Soc., 432, 1483

Carrasco Kind M., Brunner R. J., 2014, arXiv:1403.0044v1

Cautun M., Cai Y.-C., Frenk C. S., 2015, arXiv:1509.00010

Ceccarelli L., Padilla N. D., Valotto C., Lambas D. G., 2006, Mon. Not. R. Astron. Soc., 373, 1440

Ceccarelli L., Paz D., Lares M., Padilla N., Lambas D. G., 2013, Mon. Not. R. Astron. Soc., 434, 1435

Ceccarelli L., Ruiz A. N., Lares M., Paz D. J., Maldonado V. E., Luparello H. E., Lambas D. G., 2015, arXiv:1511.06741

Chan K. C., Hamaus N., Desjacques V., 2014, Phys. Rev. D, 90, 103521

Chang C. et al., 2015, Phys. Rev. Lett., 115, 051301

Clampitt J., Cai Y.-C., Li B., 2013, Mon. Not. R. Astron. Soc., 431, 749

Clampitt J., Jain B., 2015, Mon. Not. R. Astron. Soc., 454, 3357

Clampitt J., Jain B., Sánchez C., 2016, Mon. Not. R. Astron. Soc., 456, 4425

Clampitt J. et al., 2016, arXiv:1603.05790

Coe D., Benítez N., Sánchez S. F., Jee M., Bouwens R., Ford H., 2006, Astron. J., 132, 926

Colberg J. M. et al., 2008, Mon. Not. R. Astron. Soc., 387, 933

Colberg J. M., Sheth R. K., Diaferio A., Gao L., Yoshida N., 2005, Mon. Not. R. Astron. Soc., 360, 216

Coleman G. D., Wu C.-C., Weedman D. W., 1980, Astrophys. J. Suppl. Ser., 43, 393

Colless M. et al., 2001, Mon. Not. R. Astron. Soc., 328, 1039

Collister A. A., Lahav O., 2004, Publ. Astron. Soc. Pacific, 116, 345

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, Astrophys. J., 647, 201

Cooper M. C. et al., 2012, Mon. Not. R. Astron. Soc., 425, 2116

Coupon J. et al., 2009, Astron. Astrophys., 500, 981

Crocce M. et al., 2016, Mon. Not. R. Astron. Soc., 455, 4301

Crocce M., Gaztañaga E., Cabré A., Carnero A., Sánchez E., 2011, Mon. Not. R. Astron. Soc., 417, 2577

Crocce M., Pueblas S., Scoccimarro R., 2006, Mon. Not. R. Astron. Soc., 373, 369

Cunha C. E., Huterer D., Busha M. T., Wechsler R. H., 2012a, Mon. Not. R. Astron. Soc., 423, 909

Cunha C. E., Huterer D., Lin H., Busha M. T., Wechsler R. H., 2012b, arXiv:1207.3347v1

Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, Mon. Not. R. Astron. Soc., 396, 2379

Dahlen T. et al., 2013, Astrophys. J., 775, 93

Dark Energy Survey Collaboration et al., 2016, Mon. Not. R. Astron. Soc., stw641

Dawson K. S. et al., 2013, Astron. J., 145, 10

de Jong J. T. A., Kuijken K., Applegate D., Begeman K., Belikov A., al E., 2013, ESO Messenger, 154, 44

De Lucia G., Blaizot J., 2007, Mon. Not. R. Astron. Soc., 375, 2

Desai S. et al., 2012, Astrophys. J., 757, 83

Diehl H. T. et al., 2014, in Proc. SPIE, Peck A. B., Benn C. R., Seaman R. L., eds., p. 91490V

Diehl H. T. et al., 2012, Phys. Procedia, 37, 1332

Drinkwater M. J. et al., 2010, Mon. Not. R. Astron. Soc., 401, 1429

Eisenstein D. J. et al., 2001, Astron. J., 122, 2267

Elyiv A., Marulli F., Pollina G., Baldi M., Branchini E., Cimatti A., Moscardini L., 2015, Mon. Not. R. Astron. Soc., 448, 642

Feldmann R. et al., 2006, Mon. Not. R. Astron. Soc., 372, 565

Flaugher B., 2005, Int. J. Mod. Phys., A20, 3121

Flaugher B. et al., 2015, Astron. J., 150, 150

Flaugher B. L. et al., 2012, Soc. Photo-Optical Instrum. Eng. Conf. Ser., 8446, 844611

Freund Y., Schapire R., 1997, JCSS, 55, 119

Garilli B. et al., 2008, Astron. Astrophys., 486, 683

Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, Astrophys. J., 715, 823

Glorot X., Bordes A., Bengio Y., 2011, in JMLR W&CP Proc. Fourteenth Int. Conf. Artif. Intell. Stat. (AISTATS 2011)

Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, Astrophys. J., 622, 759

Graff P., Feroz F., Hobson M., Lasenby A., 2013, ArXiv e-prints

Granett B. R., Neyrinck M. C., Szapudi I., 2008, Astrophys. J., 683, L99

Greisel N., Seitz S., Drory N., Bender R., Saglia R. P., Snigula J., 2013, Astrophys. J., 768, 117

Gruen D. et al., 2016, Mon. Not. R. Astron. Soc., 455, 3367

Hamaus N., Pisani A., Sutter P. M., Lavaux G., Escoffier S., Wandelt B. D., Weller J., 2016, arXiv:1602.01784

Hamaus N., Sutter P., Lavaux G., Wandelt B. D., 2014a, J. Cosmol. Astropart. Phys., 2014, 013

Hamaus N., Sutter P., Lavaux G., Wandelt B. D., 2015, J. Cosmol. Astropart. Phys., 2015, 036

Hamaus N., Sutter P. M., Wandelt B. D., 2014, Phys. Rev. Lett., 112, 251302

Hamaus N., Wandelt B. D., Sutter P. M., Lavaux G., Warren M. S., 2014b, Phys. Rev. Lett., 112, 041304

Heymans C. et al., 2012, Mon. Not. R. Astron. Soc., 427, 146

Hildebrandt H. et al., 2010, Astron. Astrophys., 523, A31

Hildebrandt H. et al., 2012, Mon. Not. R. Astron. Soc., 421, 2355

Hildebrandt H., Wolf C., Benítez N., 2008, Astron. Astrophys., 480, 703

Hogg D. W. et al., 1998, Astron. J., 115, 1418

Hotchkiss S., Nadathur S., Gottlober S., Iliev I. T., Knebe A., Watson W. A., Yepes G., 2014, Mon. Not. R. Astron. Soc., 446, 1321

Hoyle F., Vogeley M. S., 2002, Astrophys. J., 566, 641

Ilbert O. et al., 2006, Astron. Astrophys., 457, 841

Jarvis M. et al., 2015, arXiv:1507.05603

Jennings E., Li Y., Hu W., 2013, Mon. Not. R. Astron. Soc., 434, 2167

Kaiser N., Tonry J. L., Luppino G. A., 2000, Publ. Astron. Soc. Pacific, 112, 768

Karatzoglou A., Smola A., Hornik K., Zeileis A., 2004, J. Stat. Softw., 11, 1

Kilbinger M., 2015, Reports Prog. Phys., 78, 086901

Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, Astrophys. J., 467, 38

Kitaura F.-S. et al., 2015, arXiv:1511.04405

Kuehn K. et al., 2014, AAObserver, 125, 4

Kwan J. et al., 2016, arXiv:1604.07871

Lam T. Y., Clampitt J., Cai Y.-C., Li B., 2015, Mon. Not. R. Astron. Soc., 450, 3319

Lambas D. G., Lares M., Ceccarelli L., Ruiz A. N., Paz D. J., Maldonado V. E., Luparello H. E., 2016, Mon. Not. R. Astron. Soc. Lett., 455, L99

Lavaux G., Wandelt B. D., 2010, Mon. Not. R. Astron. Soc., 403, 1392

Lavaux G., Wandelt B. D., 2012, Astrophys. J., 754, 109

Le Fèvre O. et al., 2013, Astron. Astrophys., 559, A14

Le Fèvre O. et al., 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrum. Des. Perform. Opt. Ground-based Telesc., Iye M., Moorwood A., eds., pp. 1670–1681

Le Fèvre O. et al., 2005, Astron. Astrophys., 439, 845

Lee J., Park D., 2009, Astrophys. J., 696, L10

Leistedt B. et al., 2015, arXiv:1507.05647

Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511

Li B., 2011, Mon. Not. R. Astron. Soc., 411, 2615

Liang Y., Zhao C., Chuang C.-H., Kitaura F.-S., Tao C., 2015, arXiv:1511.04391

Lilly S. J. et al., 2007, Astrophys. J. Suppl. Ser., 172, 70

Lilly S. J. et al., 2009, Astrophys. J. Suppl. Ser., 184, 218

Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, Mon. Not. R. Astron. Soc., 390, 118

Mannucci F., Basile F., Poggianti B., Cimatti A., Daddi E., Pozzetti L., Vanzi L., 2001, Mon. Not. R. Astron. Soc., 326, 745

Mao Q., Berlind A. A., Scherrer R. J., Neyrinck M. C., Scoccimarro R., Tinker J. L., McBride C. K., Schneider D. P., 2016, arXiv:1602.06306

Martí P., Miquel R., Bauer A., Gaztañaga E., 2013, Mon. Not. R. Astron. Soc., 437, 3490

Massara E., Villaescusa-Navarro F., Viel M., Sutter P., 2015, J. Cosmol. Astropart. Phys., 2015, 018

McMahon R. G., Banerji M., Gonzalez E., Koposov S. E., Bejar V. J., Lodieu N., Rebolo R., 2013, The Messenger, 154, 35

Melchior P., Sutter P. M., Sheldon E. S., Krause E., Wandelt B. D., 2014, Mon. Not. R. Astron. Soc., 440, 2922

Mohr J. J. et al., 2012, Proc. SPIE, 8451, 8451

Nadathur S., Hotchkiss S., Diego J. M., Iliev I. T., Gottlober S., Watson W. A., Yepes G., 2015, Mon. Not. R. Astron. Soc., 449, 3997

Neyrinck M. C., 2008, Mon. Not. R. Astron. Soc., 386, 2101

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, Mon. Not. R. Astron. Soc., 396, 19

Oyaizu H., Lima M., Cunha C. E., Lin H., Frieman J., 2008a, Astrophys. J., 689, 709

Oyaizu H., Lima M., Cunha C. E., Lin H., Frieman J., Sheldon E. S., 2008b, Astrophys. J., 674, 768

Padilla N. D., Ceccarelli L., Lambas D. G., 2005, Mon. Not. R. Astron. Soc., 363, 977

Paz D., Lares M., Ceccarelli L., Padilla N., Lambas D. G., 2013, Mon. Not. R. Astron. Soc., 436, 3480

Platen E., Van De Weygaert R., Jones B. J. T., 2007, Mon. Not. R. Astron. Soc., 380, 551

Pollina G., Baldi M., Marulli F., Moscardini L., 2016, Mon. Not. R. Astron. Soc., 455, 3075

Pyle, 1999, Data Preparation for Data Mining. pp. 271–274,355–359

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, Astrophys. J., 771, 30

Rozo E. et al., 2015, arXiv:1507.05460

Ruiz A. N., Paz D. J., Lares M., Luparello H. E., Ceccarelli L., Lambas D. G., 2015, Mon. Not. R. Astron. Soc., 448, 1471

Rykoff E. S. et al., 2014, Astrophys. J., 785, 104

Sahlén M., Zubeldía Í., Silk J., 2016, Astrophys. J., 820, L7

Sánchez C. et al., 2014, Mon. Not. R. Astron. Soc., 445, 1482

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, Astrophys. J., 500, 525

Sérsic J. L., 1963, Bol. la Asoc. Argentina Astron., 6, 41

Sheldon E. S. et al., 2004, Astron. J., 127, 2544

Sheth R. K., van de Weygaert R., 2004, Mon. Not. R. Astron. Soc., 350, 517

Song H., Lee J., 2009, Astrophys. J., 701, L25

Spolyar D., Sahlén M., Silk J., 2013, Phys. Rev. Lett., 111, 241103

Springel V., 2005, Mon. Not. R. Astron. Soc., 364, 1105

Strauss M. A. et al., 2002, Astron. J., 124, 1810

Sutter P. M., Lavaux G., Hamaus N., Wandelt B. D., Weinberg D. H., Warren M. S., 2014a, Mon. Not. R. Astron. Soc., 442, 462

Sutter P. M., Lavaux G., Wandelt B. D., Weinberg D. H., 2012, Astrophys. J., 761, 44

Sutter P. M., Pisani A., Wandelt B. D., Weinberg D. H., 2014b, Mon. Not. R. Astron. Soc., 443, 2983

Sypniewski A., 2014, PhD thesis, University of Michigan

Sypniewski A., Gerdes D., 2013, ArborZ home page. \url{https://cdcvs.fnal.gov/redmine/projects/arborz}

Tipping, 2001, J. Mach. Learn. Res., 1, 211

Tyson J. A., Wittman D. M., Hennawi J. F., Spergelb D. N., 2003, Nucl. Phys. B - Proc. Suppl., 124, 21

Vanzella E. et al., 2004, Astron. Astrophys., 423, 761

Wechsler R. H., 2004, Clust. Galaxies Probes Cosmol. Struct. Galaxy Evol. from Carnegie Obs. Centen. Symp. Carnegie Obs. Astrophys. Ser. Ed. by J.S. Mulchaey, A. Dressler, A. Oemler, 2004. Pasadena Carnegie Ob

Yang L. F., Neyrinck M. C., Aragón-Calvo M. A., Falck B., Silk J., 2015, Mon. Not. R. Astron. Soc., 451, 3606

York D. G. et al., 2000, Astron. J., 120, 1579

Zhao C., Tao C., Liang Y., Kitaura F.-S., Chuang C.-H., 2015, arXiv:1511.04299

Zivick P., Sutter P. M., Wandelt B. D., Li B., Lam T. Y., 2015, Mon. Not. R. Astron. Soc., 451, 4215

Anderson, L, et al. 2014, MNRAS, 441, 24

Baldauf, T., Smith, R. E., Seljak, U., & Mandelbaum, R. 2010, Phys. Rev. D, 81, 063531

Becker, M. R., Troxel, M. A., MacCrann, N., et al. 2015, arXiv:1507.05598

Behroozi, P. S., Wechsler, R. H. & Wu, H.-Y., 2013, ApJ, 762, 109

Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., Busha, M. T., Klypin, A. A. & Primack, J. R., 2013, ApJ, 763, 18

Benítez, N. 2000, ApJ, 536, 571 Data Analysis Software and Systems XX, 442, 435

Betoule, M, et al. 2014, A&A, 568, 22

Beutler, F., Blake, C., Colless, M., Jones, D. H., Staveley-Smith, L., Campbell, L., Parker, Q, Saunders, W, et al., 2011, MNRAS, 416, 3017

Blanton, M. R., Hogg, D. W., Bahcall, N. A., Brinkmann, J., Britton, M., Connolly, A. J., Csabai, I., Fukugita, M., et al., 2003, ApJ, 592, 819

Blazek, J., McQuinn, M., & Seljak, U., 2011, JCAP, 5, 1

Blazek, J., Mandelbaum, R., Seljak, U., & Nakajima, R, 2012, JCAP, 5, 41

Blazek, J., Vlah, Z., & Seljak, U., 2015, JCAP, 8, 15

Bonnett, C., Troxel, M. A., Hartley, W., et al. 2015, arXiv:1507.05909

Bonnett, C., 2015, MNRAS, 449, 1043

Brainerd, T. G., Blandford, R. D., Smail, I., 1996, ApJ, 466, 623

Carrasco Kind, M., & Brunner, R., J., 2013, MNRAS, 432, 1483

Cacciato, M., van den Bosch, F. C., More, S., et al. 2009, MNRAS, 394, 929

Cacciato, M., van den Bosch, F. C., More, S., Mo, H., & Yang, X. 2013, MNRAS, 430, 767

Catelan, P., Kamionkowski, M., & Blandford, R. D., 2001, MNRAS, 320, 7

Chang, C., Vikram, V., Jain, B., Bacon, D., Amara, A., Becker, M. R., Bernstein, G. & Bonnett, C. et al., 2015, Physical Review Letters, 115, 1301

Cooray, A., & Sheth, R. 2002, PhR, 372, 1

Clampitt, J., Sánchez, C., Kwan, J., et al. 2016, arXiv:1603.05790

Crocce, M., Carretero, J., Bauer, A. H., et al. 2015, arXiv:1507.05360

Cunha, C. E., Huterer, D., Busha, M. T. & Wechsler, R. H., 2012, MNRAS, 423, 909

The Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2015, arXiv:1507.05552

The Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, arXiv:1601.00329

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25

de Jong, J. T. A., Verdoes Kleijn, G., A., Boxhoorn, D. R., Buddelmeijer, H., Capaccioli, M., Getman, F., Grado, F., Helmich, E., et al., 2015, A&A, 582, A62

Efstathiou, G, 2014, MNRAS, 440, 1138

Erben, T., Hildebrandt, H., Miller, L., et al. 2013, MNRAS, 433, 2545

Feroz, F., Hobson, M. P., & Bridges, M., 2009,MNRAS, 398, 1601

Flaugher, B. L., Abbott, T. M. C., Angstadt, R., et al. 2012, SPIE, 8446, 844611

Flaugher, B., Diehl, H. T., Honscheid, K., Abbott, T. M. C., Alvarez, O., Angstadt, R., Annis, J. T., Antonik, M. et. al, 2015, AJ 150, 150.

Friedrich, O., Seitz, S., Eifler, T. F. & Gruen, D., 2015, arXiv:1508.00895

Giannantonio, T., Fosalba, P. et al. 2015, arXiv:1507.05551

Gillis, B. R., Hudson, M. J., Erben, T., et al. 2013, MNRAS, 431, 1439

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Graff, P., Feroz, F., Hobson, M. P., & Lasenby, A. 2014, MNRAS, 441, 1741

Gruen, D., Friedrich, O., Amara, A., et al. 2015, arXiv:1507.05090

Guzik, J. & Seljak, U, 2001, MNRAS, 321, 439

Heymans, C., Van Waerbeke, L., Miller, L., Erben, T., Hildebrandt, H., Hoekstra, Kitching, T. D., et al., 2012, MNRAS, 427, 146

Heymans, C., Grocutt, E., Heavens, A., Kilbinger, M., Kitching, T D., Simpson, F., Benjamin, J., Erben, T. et al, 2013 MNRAS, 432, 2433

Hartlap, J., Simon, P. & Schneider, P. 2007, A&A, 464, 399

Hirata, C. & Seljak, U., 2004, Phys. Rev. D, 70, 063526

Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, JCAP, 4, 027

Hudson, M. J., Gillis, B. R., Coupon, J., et al. 2015, MNRAS, 447, 298

Jarvis, R. M., Bernstein, G.,& Jain, B., 2004, MNRAS, 352 338

Jarvis, M., Sheldon, E., Zuntz, J., et al. 2015, arXiv:1507.05603

Joachimi, B., Mandelbaum, R., Abdalla, F. B. & Bridle, S. L., 2011, A&A, 527, 26

Joachimi, B., Cacciato, M., Kitching, T. D., Leonard, A., Mandelbaum, R., Schäfer, B. M., Sifón, C. Hoekstra, H., et al., Space Science Reviews, 193, 1

Kaiser, N., 1992, ApJ, 388, 272,

Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500

Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64

Lawrence, E., Heitmann, K., White, M., Higdon, D., Wagner, C., Habib, S., & Williams, B., 2010, ApJ, 713, 1322

Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159

Lehmann, B. V., Mao, Y.-Y., Becker, M. R., Skillman, S. W. & Wechsler, R. H. 2015, arXiv:1510.05651

Leistedt, B., Peiris, H. V., Elsner, F., et al. 2015, arXiv:1507.05647

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Li, T. S., DePoy, D. L., Marshall, J. L., et al. 2016, arXiv:1601.00117

Limber, D., N., 1953, ApJ, 117, 134

Mandelbaum, R., Seljak, U., Cool, R. J., et al. 2006, MNRAS, 372, 758

Mandelbaum, R., Seljak, U., & Hirata, C. M. 2008, JCAP, 8, 6

Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, MNRAS, 432, 1544

Mao, Y-Y, Williamson, M., & Wechsler, R. H., 2015, ApJ, 810, 21

Marian, L, Smith, R. E. & Angulo, R. E., 2015, MNRAS, 451, 1418

McCarthy, I. G., Schaye, J., Bower R., G., Ponman T. J., Booth, C. M., Dalla Vecchia, C., & Springel, V., 2011, MNRAS, 42, 1965

McDonald, P., 2006, Phys. Rev. D, 74, 103512

Miyazaki, S., Komiyama, Y., Nakaya, H., Kamata, Y., Doi, Y., Hamana, T., Karoji, H., Furusawa, H., et al., 2012, SPIE, 8446, 0Z

Monet, D. G., Levine, S. E., Canzian, B., Ables, H. D., et al. 2003, AJ, 125, 984

More, S., Miyatake, H., Mandelbaum, R., et al. 2015, ApJ, 806, 2

Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, MNRAS, 396, 19

Park, Y., Krause, E., Dodelson, S., et al. 2015, arXiv:1507.05353

The Planck Collaboration, Ade., P. A. R et al, 2015, arXiv:1502.01589

Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S., 2013, ApJ, 771, 30

Ross, A., J. et al. 2011, MNRAS, 417, 1350

Ross, A. J. et al. 2012, MNRAS, 424, 564

Ross, A. J. et al. 2014, MNRAS, 437, 1109

Ross, A. J. et al. 2015, MNRAS, 449, 835

Rozo, E., Rykoff, E. S., Becker, M., Reddick, R. M., & Wechsler, R. H. 2015, MNRAS, 453, 38

Rozo, E., Rykoff, E. S., Abate, A., et al. 2015, arXiv:1507.05460

Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104

Sadeh I., Abdalla F., B., Lahav O., 2015, arXiv:1507.00490

Seljak U., 2000, MNRAS, 318, 203

Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, MNRAS, 445, 1482

Sheldon, E. S., Johnston, D. E., Frieman, J. A., et al. 2004, AJ, 127, 2544

Sheldon, E. S. 2014, MNRAS, 444, L25

Sifón, C., Cacciato, M., Hoekstra, H., et al. 2015, MNRAS, 454, 3938

Singh, S., & Mandelbaum, R., 2015, MNRAS, 450, 2195

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311

Springel, V., 2005, MNRAS, 364, 1105

Suchyta, E., Huff, E. M., Aleksic, J., Melchior, P., et al. 2016, MNRAS, 457, 786

Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152

Troxel, M. A. & Ishak, M., 2015, PhR, 558, 1T

Tyson, J. A., Valdes, F., J. F., & Mills, A. P., Jr. 1984, ApJL, 281, L59

van Daalen, M. P., Schaye J., Booth C. M. & Dall Vecchia C., 2011, MNRAS, 415, 3649

van Uitert, E., Cacciato, M., Hoekstra, H., et al. 2016, arXiv:1601.06791

Velander, M., van Uitert, E., Hoekstra, H., et al. 2014, MNRAS, 437, 2111

Vikram, V., Chang, C., Jain, B., et al. 2015, Phys. Rev. D, 92, 022006

Viola, M., Cacciato, M., Brouwer, M., et al. 2015, MNRAS, 452, 3529

Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2006, MNRAS, 373, 1159

Yoo, J., & Seljak, U. 2012, Phys. Rev. D, 86, 083504

York, D. G., Adelman, J., Anderson, J. E., Jr., Anderson, S. F., Annis, J., Bahcall, N. A., Bakken, J. A., Barkhouser, R., et al., 2000, AJ, 120, 9

Zuntz, J., Kacprzak, T., Voigt, L., et al. 2013, MNRAS, 434, 1604

Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45