A Bayesian Application of Redshift Distributions to Weak lensing Surveys

George Kyriacou

Department of Physics
Imperial College London

A thesis submitted for the degree of
Doctor of Philosophy

July 2023

Abstract

As we move into the next era of cosmology, dominated by large extensive surveys in the name of precision. The question of how best to constrain our current models with the vast amount of data available to us now and in the future has become prominent. This thesis looks to answer that question in the field of weak lensing, specifically for redshift distributions, a typical weakness of the field.

To begin, this thesis will look at the cosmological parameters that describe our universe to the best of our ability and the probes available for constraining them. Chapter 2 looks at one of those probes in more detail, weak lensing, and how a light bending phenomenon can be used to infer statistical properties of the matter distribution of the universe and how its precision is hampered by weaknesses in redshift inference. Methods for achieving this is the focus of Chapter 3, photometric redshifts; using light collected from galaxies across a small number of broadband filters to infer its redshift based on known spectral features. We look into how such methods developed, how these spectral features exist, and the various approaches used. One approach requires a change to our statistical thinking, and thus Chapter 4 focuses on Bayesian statistics and ways it can be applied to problems. One such application is considered in Chapter 5 where we modify the work of Leistedt et al. (2016) for weak lensing-like settings. In Chapter 6 we apply the work to Kilo Degree Survey, specifically KV450 and constrain cosmological parameters potentially lowering a current tension in S_8 values between weak lensing and CMB studies. Lastly, in Chapter 7, we gather preliminary results using our approach to KiDS-1000 data, considering improvements required for the levels of precision and accuracy of future surveys.

Declaration

I hereby declare that the contents of this thesis are my own work, except where specific reference is made to the work of others. The specific contributions in each chapter are listed below.

Chapters 1, 2 and 3 correspond to reviews and general work on the past and present of precision cosmology, weak lensing and photometric redshift estimation. In particular, in the reviews of Kilbinger (2014) and Mandelbaum (2017). All sources of information are appropriately referenced.

Chapter 4 corresponds to work and reviews on Bayesian statistics and a key focus on the works of Malz and Hogg (2020) in redshift distribution inference.

Chapters 5, 6 and 7 describe my work in collaboration with Alan Heavens, Andrew Jaffe, Florent Leclercq and Arrykrishna Mootoovaloo, which expands upon and is heavily based on the work of Leistedt et al. (2016)

George Kyriacou
July 2023

Acknowledgements

I want to thank my supervisors Alan Heavens, Andrew Jaffe and Florent Leclercq, for the years of guidance and patience they have given me. The many months of frustration, circling, stop and start, re-explaining and sheer time and effort that's a consequence of working with me. A special thanks to Harry (Arrykrishna Mootoovaloo) for his kindness and generosity in all the voluntary help he has given me, especially near the end. To my mum (Hilary Kyriacou) for taking on the hellish task of painstakingly proofreading this thesis for my many grammatical and spelling errors. Most importantly, I would like to thank my wife, Shannon Kyriacou, for the sacrifices she has made for me to achieve my goals and ambitions and for never wavering in her support of me during a PhD of which, at times, I could barely support myself.

Contents

List of figures

List of tables

Chapter 1

Cosmological Parameters

If one was to describe the state of modern cosmology, one might start with the brief term 'precision cosmology'. However, with slightly more details, these words unpack to be the study of statistical techniques to improve the inference of cosmology parameters given observational data and their chosen model. This statement carries with it the many aspects this thesis will cover. First, the observational data is the survey images of the sky, giving us both colour and shape information of millions of galaxies. The chosen λCDM model and its parameters describe our universe and how this relates to the galaxy shape in terms of weak lensing. The inference relates to the tools used to link both model and data to their likely parameters correctly. Lastly, the statistical techniques, the main focus of this thesis, look into how one can correctly infer survey redshift distributions from flux data. On the surface, this may look like a small cog in a vast cosmological machine, but a cog can have detrimental effects if not treated correctly. To adequately justify the impact this small statistical conundrum can have on our understanding of cosmology, one has, to begin with, that cosmology itself, the assumptions it makes, the variables it holds and the probes used to investigate it.

1.1 Einstein's Equations

Although aptly described as 'the weakest fundamental force' (Shirai and Yamazaki, 2019) compared to its stronger counterparts (electromagnetic, strong, weak), the large masses of planets, stars, galaxies and galaxy clusters and the large distances between them make gravity the overwhelming force of the universe. Therefore, gravity is the key to understanding the past and present and future on a macro scale.

Huge strides were made in the revolution that was 'The theory of General Relativity (Einstein et al., 1915) (for now on summarised as GR) and the jumping of the gravitational

domain from Newton to Einstein. As a result, gravity was no longer a force instantaneously reacting to surrounding matter density (already a violation of Special Relativity's strict speed limits (Einstein, 1905)) but instead the consequence of the relationship between matter and curvature in our universe and shown in Einstein's field equations:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}, \tag{1.1}$$

where $\mu\nu$ are index values, $g_{\mu\nu}$ is the metric tensor, $R_{\mu\nu}$ and R are the Ricci Tensor and curvature scalar, respectively, G the gravitational constant and $T_{\mu\nu}$ the energy-momentum tensor. Each of these terms helps describe the matter-curvature relationship mentioned above. The metric tensor describes the coordinate system used to give an infinitesimal space-time step ds along the system.

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \tag{1.2}$$

where dx^μ is an infinitesimal step in the μ^{th} direction. The Ricci Tensor and curvature scalar describe the coordinate system's shape. While the energy-momentum tensor describes the matter-energy distribution as a perfect fluid which in its inertial reference frame is given by

$$T_{\mu\nu} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{bmatrix}, \tag{1.3}$$

where ρ is the total energy density of the system (including the matter), and p is the isotropic pressure. Lastly, the Gravitational Constant is the same as that defined originally by Newton so that on earth-like scales, there is no disagreement with Newtonian gravity. Depending on which way round (left-to-right or right-to-left) the equation is read, one can interpret the equation as the energy-matter of the system determining the system coordinate curvature *or* the system coordinate curvature dictating the dynamics of its matter and energy. Succinctly matter is not attracted to matter due to force more that the matter is following its natural path through space-time, a natural path in which matter's presence has altered.

The most significant assumption of our current cosmological models is that GR dictates their dynamics. Such an assumption can be made as the theory has successfully guided our understanding of the universe with tests that have continued to prove its resilience for a decade. The earliest is the fact that GR could explain away observations that disagreed with Newtonian physics, such as the procession of Mercury (Treschman, 2014); the value predicted by Einstein et al. (1915) not only agreed with the observation at the time (Le Verrier

et al., 1859) but was still in agreement with recent constrained observations (Lo et al., 2013). Whereas predictions made by GR (such as Gravitational Waves (Einstein and Rosen, 1937)) continue to be confirmed in our modern era of observation (Abbott et al., 2016).

As successful as the theory continues to be, one cannot ignore its limitations. For example, the mathematics of GR predicts singularities within a black hole, indicating a point at which the theory breaks down (Einstein and Cohen, 1956). The same can be said because GR is inherently classical and continuous and has not been successfully quantised to agree with quantum field theory (Wagh, 2004). As real as these flaws may be, they exist at the extremes of our understanding, with enormous masses on the most minor scales, scales that will not be probed for the remainder of this thesis and thus, GR is a fair assumption to take forward.

1.2 Cosmological Principle

The field equation 1.1 may well be used for the dynamics of the universe but solving the differential equations is not an easy process and, at the very least, requires a metric of choice. Thankfully there are more key assumptions one can use. In this case, the following fundamental assumption is the Cosmological Principle. The principle states that "the universe is isotropic and homogeneous, i.e. there is no preferred direction or preferred position" (Piattella, 2018).

Originally the principle was stated without proof as a means to simplify models (Lahav, 2001). This is because the principle at its heart is challenging to verify (all more likely falsified). Furthermore, as we are earthly bound, one can only observe the universe from one point in space. These limits are observations within our light cone (objects outside the light cone have not had enough time for their light to reach us). However, homogeneity is observable within our available scales, such as the distribution of quasars, which were observed to be homogeneous on scales larger than $250\,h^{-1}$ Mpc (Sarkar and Pandey, 2016). In contrast, the isotropy is much more easily observed (due to all angles being available for observation), with all models of anisotropy being highly constrained. Furthermore, although homogeneity and isotropy are independent quantities, one can use validation isotropy to constrain homogeneity further as long as one concedes a much smaller assumption; the Copernican principle (Maartens, 2011) which states that we are not in a unique position in the universe (Peacock, 1998).

There are more formal definitions of the cosmological principle that help construct a useful metric which leads to 2 conditions below (Weinberg and Wagoner, 1973):

1. the constant-time hypersurfaces are maximally symmetric

2. matter quantities are only time-dependent.

The first condition means that a maximally symmetric space is completely characterised by 1 constant number, the scalar curvature R whereas the Ricci Tensor summarises to $R_{\mu\nu} = g_{\mu\nu}R/D$ where $D = g^{\mu\nu}g_{\mu\nu}$. As R can be negative, positive or zero, this leaves 3 possible 3-dimensional special cases.

1. $R = 0$ therefore $ds_3^2 = |dx|^2$ a flat euclidean space,
2. $R > 0$ therefore $ds_3^2 = |dx|^2 + dz^2$ under the condition that $z^2 + |x^2| = a$ where a is a constant. This is a euclidean space wrapped around a sphere of radius a.
3. $R < 0$ therefore $ds_3^2 = |dx|^2 - dz^2$ under the condition that $z^2 - |x^2| = a$ where a is a constant. This is a euclidean space on a hyperboloid.

These 3 possibilities are usually known as the universe being 'flat', 'closed' or 'open', although all 3 are possible under the cosmological principle. Observations indicate a flat universe (Lahav and Liddle, 2014).

Moving to spherical coordinates, these 3 possibilities are compact to give equation 1.4.

$$ds_3^2 = a^2\left(\frac{dr^2}{1-Kr^2} + r^2 d\Omega^2\right), \tag{1.4}$$

where Ω is the solid angle, r is the normalised radius and $K = R^{(3)}a^2/6$ so that the 3 possibilities are determined by K are 0,-1 or 1. To complete the 4D space-time metric, one must incorporate time. As the CP definitions state, the special components can still scale (and thus be variable) with time as long as they remain symmetric at a constant time and thus were allowed to make a a function of time, producing the Friedmann-Robertson-Walker metric (Friedman, 1922; Friedmann, 1924; Lematre and Eddington, 1931; Robertson, 1935, 1936):

$$ds^2 = -dt^2 + a^2(t)(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2), \tag{1.5}$$

where t is the cosmic time, such a metric will make calculations within GR easier and more tethered to our understanding. Furthermore, recent numerical relativity simulations have shown that generic metrics are compatible with FRW metrics as long as homogeneity persists on large scales; thus, its use is justified (Giblin et al., 2016). However, this opens a new door regarding our scale factor $a(t)$. In what sense is our universe spatially time-dependent? In what sense does our universe scale?

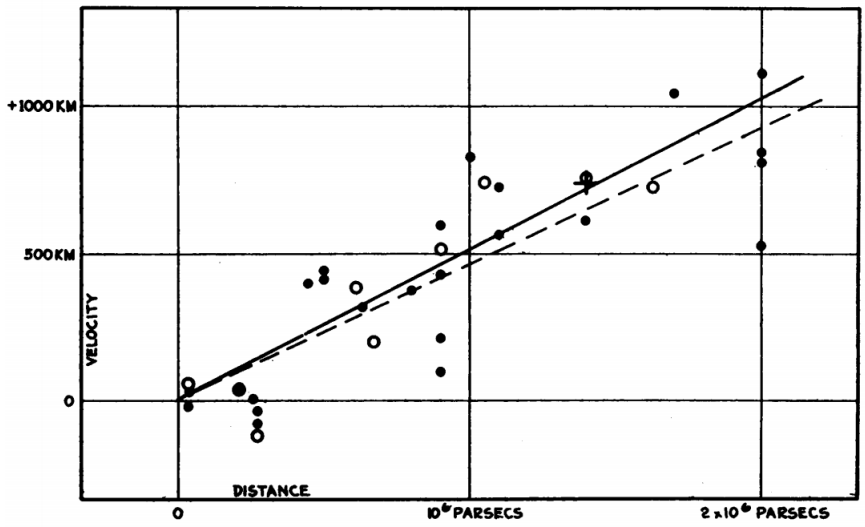


Figure 1.1: Radial Velocities corrected for solar motion are plotted against distances from stars and the mean luminosity of the nebula in a cluster. Black disks and the full line represent the use of nebulae individually, while circles and the dashed line represent combining nebular into groups. The cross is the mean velocity and mean distance of 22 nebular that could not be estimated individually (Hubble, 1929).

1.3 Expanding universe

As stated above, the FRW metric implies a homogeneous, isotropic non-static universe. One that may look the same from every position and direction but not necessarily look the same from different periods in time. The nature of this time-based dependency is entirely one way; towards expansion. The first evidence of cosmological expansion came from observations that galaxies that are further away are moving faster away from us. This is most commonly known as Hubble's Law:

$$v = H_0 d, \tag{1.6}$$

where v is the galaxy's velocity away from us, d is the distance that galaxy is from us, and H_0 is the constant relating the two (known as the Hubble Constant). Figure 1.1 shows the original data used to which Hubble observed the first value of H_0 to be $H_0 = 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Hubble, 1929), now known to be systematically wrong.

This original value of H_0 was highly imprecise due to a large amount of scattering in the original data. However modern observations have since constrained H_0 to be lie between $66 - 77\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Riess, 2020). That said, independent observations of H_0 have to lead to values that are no longer compatible with each other (an issue known as 'Hubble Tension' (Efstathiou, 2021). The most notable examples of said tension are the SH0ES (Supernovae H_0 for the Equation of State) collaboration which uses a distance ladder approach. The

approach uses the luminosity of specific star types (Cepheids and Type 1a supernovae) as standard candles (objects of known luminosity) to establish distance measures. This gave a value of $H_0 = 74.03 \pm 1.42\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Riess et al., 2019) and observations of the Cosmic Microwave Background obtained a value of $H_0 = 67.44 \pm 0.58\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Collaboration et al., 2018) which is a disagreement in value by over 4σ. Much research has taken place to discover if the tension is theoretical, showing the need for modifications of our current cosmological models to be resolved (Knox and Millea, 2019). Alternatively, if the tension is purely statistical and results from the data not being analysed correctly while the underlying theory stays intact (Freedman et al., 2020), for this thesis, the tension (and implications) will be mostly ignored under the assumption that the true value exists within the above range with little change to our current models. One can take confidence from this assumption as many other independent methods of measuring H_0 all agree that H_0 exists within the said range (Shajib et al., 2019; Wong et al., 2019).

1.3.1 Redshift

A direct consequence of an expanding universe would be that the light we receive from distant galaxies would be 'redshifted'. By using Equation 1.5 as the metric within Equation 1.1, one can show that the energy of a photon is inversely proportional to the scale factor; thus, one can link the scale factor directly to the wavelength of observed light:

$$\frac{a_{\mathrm{em}}}{a_{\mathrm{obs}}} = \frac{E_{\mathrm{obs}}}{E_{\mathrm{em}}} = \frac{f_{\mathrm{obs}}}{f_{\mathrm{em}}} = \frac{\lambda_{\mathrm{em}}}{\lambda_{\mathrm{obs}}} = \frac{1}{1+z}, \tag{1.7}$$

where a, E, f, λ is the scale factor, the energy of the photon, the frequency of the photon and the wavelength of the photon at the time of emission ($_{\mathrm{em}}$) or observation ($_{\mathrm{obs}}$) while z is the redshift. By defining $a_{obs} = 1$ this simplifies to $1 + z = 1/a_{em}$.

The inclusion of redshift as the first physical consequence of cosmological expansion (outside of recession velocities of the original discoveries) brings about the question of how best to interpret the expansion going forward or, more importantly, the limitations of these interpretations. The standard interpretation is the expanding space interpretation (Pössel, 2020) that galaxies within the universe are at rest (peculiar velocities aside) and that it is the space around that is expanding. The interpretation is usually coupled with galaxies on the surface of a slowly expanding balloon, highlighting the $d \propto r$ relation of Hubble's law. However, as helpful as this interpretation can be, there are limitations to concepts such as Hubble's law or the understanding of redshift as the expansion of light waves. Firstly, it tethers our understanding of the recession velocity of galaxies to the change in a scale factor at a fixed time. This leads to the intuitive but untrue assumption that in an expanding

universe, a galaxy that was momentarily at rest with us must eventually move away due to the space expanding between us (in fact, even in an expanding universe, the galaxy may move towards us as highlighted in the Tethered Galaxy Problem (Francis et al., 2007)). Also, with redshift, one may incorrectly imagine the electromagnetic waves of the observed photons to be stretching.

An alternative interpretation is that space itself does not expand. It is the galaxies themselves moving through space. Under such rules, redshift would be observed as a Doppler shift. This method also has faults, if one was to work out the radial velocity required for observed redshifted galaxies, it would not equal the recession velocity. Instead, a relativistic radial velocity is required for the redshift to match as below:

$$z = \sqrt{\frac{c + v_{\text{rad}}}{c - v_{\text{rad}}}} - 1, \tag{1.8}$$

where c is the speed of light and v_{rad} is the relativistic radial velocity (Narlikar, 1994). Even with this change, the interpretation would fail to explain the counter-intuitive results that galaxies at a constant distance from us can have a non-zero redshift (Davis et al., 2003).

As shown, both interpretations can be mathematically equivalent if conceptually limited. For this thesis, the expansion will be referred to in more modest words as the consequence of 2 bodies existing within the FRW metric. Redshift is the consequence of an observer's frame of reference being different to that of the photon's emitter due to the FRW metric and not a continuous change to the photon. Furthermore, this eliminates looming questions such as "Is the space in a bedroom expanding?" (J.A Peacock et al., 2001) as a bedroom is not within an FRW metric (the matter that makes that room neither homogeneous nor isotropic).

Lastly, although not a focus of this thesis, the redshifting of galaxies is not exclusively due to the universe's expansion. General relativity also predicts the existence of redshift due to gravitation as shown in the equation 1.9 and is often known as Einstein shift (Eddington, 1926):

$$z_g = \frac{\Delta\phi}{c^2}, \tag{1.9}$$

where $\Delta\phi$ is the change in gravitational potential. This gravitational effect has been verified both in the lab (Pound and Rebka, 1960) and observed in space (Adams, 1925). One can even correct the equation1.8 to include gravitational redshift as seen in the equation 1.10 (Peacock, 2008).

$$1 + z = \sqrt{\frac{1 + v_{\text{rad}}/c}{1 - v_{\text{rad}}/c}}(1 + \frac{\Delta\phi}{c^2}). \tag{1.10}$$

Previous studies of gravitational redshift have used data from the Sloan Digital Survey (Eisenstein and Weinberg, 2011). By comparing galaxies at the centre of galaxy clusters (usually in the form of brightest cluster galaxies) with galaxies on the outskirts of clusters (to have similar z_H with varying gravitational potentials), gravitational redshifts were estimated to be of the order of $z_g \sim 10^{-5}$ (Wojtak et al., 2011)(Sadeh et al., 2014) (or a velocity contribution of $\sim 10\,\mathrm{km\,s^{-1}}$). An effect minimal enough to be ignored for the large redshifts being considered in this thesis. Galaxies also have their peculiar velocities v_{pec}, which is moving relative to us and differs from the cosmic expansion or 'Hubble Flow'. As this velocity still contributes towards the overrule change of frame between emitter and observer, it also produces its redshift z_{pec} equivalent to the equation 1.8 but with v_{pec} instead of v_{rad}. At low redshift, this is approximated to $v_{\mathrm{pec}} \approx c z_{\mathrm{pec}}$ allowing the following combination for the observed redshift (Davis et al., 2010),

$$1 + z_{\mathrm{obs}} = (1 + z_H)(1 + z_g)(1 + z_{\mathrm{pec}}), \tag{1.11}$$

where z_H, z_g, and z_{pec} are the redshifts due to the Hubble flow, Einstein shift and peculiar velocities, respectively.

Isolating peculiar velocities from the expansion velocities requires knowledge of the radial velocity relative to a reference system and an independent method of measuring galaxy distance. The latter usually comes in the form of the Tully-Fisher relation (Tully and Fisher, 1977) (that luminosity of a spiral galaxy follows a power law with rotation velocity), which was used by the SFI++ (Masters et al., 2006) and 2MTF (Masters et al., 2008) catalogues. The fundamental plane relation (Dressler et al., 1987) (that the radius of a galaxy is closely related to its surface brightness and dispersion velocity) used by the 6dF velocity survey (Springob et al., 2014). An alternative approach reconstructs and simulates the expected velocity field and compares it to the observed galaxies (Lavaux et al., 2010)(Erdogdu et al., 2006). Analysis of galaxy clusters finds average peculiar velocities of $\sim 300\,\mathrm{km\,s^{-1}}$ (Guidorzi et al., 2017)(Karachentsev et al., 2006)(Boruah et al., 2019)(Nicolaou et al., 2019) which approximates to a redshift contribution of $z_{\mathrm{pec}} \approx 0.001$ and so just as with Einstein shift above z_{pec} is also presumed to be low in comparison to z_H and ignored for future calculations.

1.3.2 Cosmological Constant

The discovery of an expanding universe was met with the assumption that due to gravity, galaxies would slowly pull back closer and that the expansion would eventually stop. So confident was this assumption that the 'deceleration parameter' q_0 was introduced into cosmology, believing it to be positive (Sandage, 1970). However, 2 independent analyses

of type 1a supernova and their distance-redshift revealed that q_0 would be negative, and the expansion was accelerating (Riess et al., 1998)(Perlmutter et al., 1998). The analysis can be summarised as the supernova redshifts observed were less than expected given the supernova's known distance from us. By 'expected', one means redshift growing slower with distance than the linear relationship one would get from constant expansion and much slower than with decelerating expansion. Met initially with scepticism, proof that the universe is accelerating in its expansion has now been observed independently through Cosmic Microwave Background (Clarkson, 2012), and Baryon Acoustic Oscillations (Eisenstein et al., 2005).

Accelerated expansion can be mathematically explained with modifications to equations 1.1 but the mathematics originally existed to solve a different problem based on incorrect assumptions. While Einstein was developing his general theory of relativity, he presumed we lived within a static universe (Einstein, 1917) (no expansion), a presumption based on observations at the time and a presumption that would not be undermined for another 12 years (Hubble, 1929). He also presumed we lived in a closed universe (positive curvature) (O'Raifeartaigh et al., 2017). Equation 1.1, when applied to FRW metric, leaves these 2 assumptions incompatible, and the closed universe would expand. A constant was added to the equation 1.1 known as the cosmological constant to remedy this conflict,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.12}$$

where Λ was the cosmological constant. Although the discoveries of Hubble would leave Einstein's static presumption redundant, the constant would still have a place in modern cosmology, especially since it allows more flexibility in a universe's expansion rate previously available.

To see this in more detail, we shall apply the metric of equation 1.5 to the revised field equations 1.12. Beginning with curvature, one can compute both the Ricci Tensor $R_{\mu\nu}$ and scalar R to our homogeneous and isotropic universe:

$$R_{00} = -3\frac{\ddot{a}}{a}, \tag{1.13}$$

$$R_{0i} = 0, \tag{1.14}$$

$$R_{ij} = \frac{1}{c^2}g_{ij}\left(\frac{\ddot{a}}{a} + \left(2\frac{\dot{a}}{a}\right)^2 + 2\frac{kc^2}{a^2}\right), \tag{1.15}$$

$$R = \frac{6}{c^2}\left(\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2}\right), \tag{1.16}$$

the dots (ie $\dot{a}$, $\ddot{a}$) represent first and second time derivatives. This helps simplify the Einstein equations,

$$H^2 + \frac{kc^2}{a^2} = \frac{8\pi G}{3c^2}T_{00} + \frac{\Lambda c^2}{3}, \tag{1.17}$$

$$g_{ij}(H^2 + 2\frac{\ddot{a}}{a} + \frac{kc^2}{a^2} - \Lambda c^2) = -\frac{8\pi G}{c^2}T_{ij}, \tag{1.18}$$

where $H = \dot{a}/a$ is the Hubble parameter producing a normalised rate of expansion. For the stress-energy tensor, we presume matter works as a perfect fluid (Schutz, 1985), so it has the following form:

$$T_{\mu\nu} = (\rho + \frac{P}{c^2})u_\mu u_\nu + P g_{\mu\nu}, \tag{1.19}$$

where u_μ is the 4-velocity of the fluid state. Perfect fluids are a straightforward version of a fluid state; they are a state in which there are no forces between particles, no heat conduction and no viscosity. One may use the perfect fluid state when approximating the universe on large scales. As we were presuming GR gravity is not considered a force, so on large scales, all other forces are negligible. This is also the same for heat conduction and viscosity, although still existing on small scales are negligible at large (Hobson et al., 2006). As forces, heat conduction and viscosity still exist, using a perfect fluid is at best a first-order approximation. Its use greatly simplifies the field equations with all non-diagonal terms going to zero, allowing for the production of equations that fit with observations (Ward, 2017).

There are criticisms of the assumption that the universe (or, more precisely, matter and energy content of the universe) can be modelled as a perfect fluid. Firstly the argument that the use of equations1.19 greatly simplifies the field equations has become mostly redundant due to efforts being made to solve FRW with more generalised fluid states (Das et al., 2017). Secondly, by restricting ourselves to the limitations of perfect fluids, one can have issues calculating non-perfect phenomena such as Bose-Einstein condensation of dark matter, heat flow and dissipative mechanisms (Comer et al., 2011). With this in mind, alternative methods of navigating the FRW equations are being developed, such as the relativistic variational multi-fluid approach (B.Carter, 1984) or many component single fluid models (Gromov et al., 2002). These approaches offer insight into cosmology's subsequent developments, but none have been convinced of their ability to match cosmological observation (Ha et al., 2012).

Going forward with the perfect fluid assumption, one can combine the stress-energy Tensor equation 1.19 with the simplified Einstein equations 1.17 and 1.18 to produce the Friedmann Equations:

$$H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2}, \tag{1.20}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + \frac{3P}{c^2}) + \frac{\Lambda c^2}{3}. \tag{1.21}$$

Considering the original problem Einstein had in which there is no cosmological constant ($\Lambda = 0$) and a closed universe $k > 0$. Under these circumstances, it's possible to have a static universe $H = 0$ from the equation 1.20 but as long as $P \geq 0$ and $\rho \geq 0$ (which is the case for all known matter at the time) then from equation 1.21 one can see that $\ddot{a} \neq 0$ thus a non-static universe is not possible. A finely tuned non-zero Λ would still allow Einstein's static universe to be compatible with both Friedmann Equations.

As stated earlier, observations strongly indicate that we do not exist within an Einstein static universe, but this discovery has not retired the constant. Once again consider a situation with no cosmological constant $\Lambda = 0$ but now we are in the antithesis of the Einstein static universe, one of which the universe is open or flat $k \leq 0$ and is accelerating in its expansion ($\dot{a} > 0, \ddot{a} > 0$). Once again under the presumption that $P \geq 0$ and $\rho \geq 0$ one is unable to prevent deceleration in the equation 1.21. However, with constant inclusion, we now have the freedom to accelerate a flat or open universe.

The rest of this thesis will presume the existence of the cosmological constant, but it is probably best to consider the Λ as the mathematical framework that allows acceleration within GR as opposed to a complete explanation. One does not even have the freedom of treating Λ as a 'quick fix' as it has physical interpretations that have led to legitimate criticisms (Solà, 2013).

This physical interpretation comes from particle physics. By modelling the universe as a scalar field, ϕ one can re-write the energy-momentum tensor regarding kinetic and potential energies.

$$T_{\mu\nu} = \frac{1}{2}\partial_\mu\phi\partial_\nu\phi + \frac{1}{2}(g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi)g_{\mu\nu} - V(\phi)g_{\mu\nu}, \tag{1.22}$$

where $\partial = \partial/\partial x_\mu$ are partial derivatives and $V(\phi)$ the potential of the field. One can consider the tensor in a vacuum where there is no kinetic energy $\partial_\mu\phi = 0$ and the potential energy has been minimised $V(\phi) = V(\phi_0)$ where ϕ_0 is the value of ϕ that gives the minimal potential. Thus the vacuum tensor can be written (Weinberg, 1989) as

$$T^{\text{vac}}_{\mu\nu} = -V(\phi_0)g_{\mu\nu} = -\rho_{\text{vac}}g_{\mu\nu}, \tag{1.23}$$

where ρ_{vac} is the vacuum energy density. One can make this equivalence in the equation 1.23 within the context of GR. This would be all that is left within the tensor. By rearranging the equation 1.12 one can consider the cosmological constant in the same manner as

$$G_{\mu\nu} = 8\pi G(T_{\mu\nu} + \rho_{\text{vac}}), \tag{1.24}$$

where $G_{\mu\nu} = R_{\mu\nu} - g_{\mu\nu}R/2 + \Lambda g_{\mu\nu}$ and $\rho_{\text{vac}} = \rho_\Lambda = \Lambda/8\pi G$. Under this rearrangement, the cosmological constant and vacuum energy are equivalent.

With this interpretation, Λ should reveal itself beyond just accelerated expansion but through other observations. One such example of vacuum energy comes from the Casimir Effect. Quantum theory suggests vacuums have zero-point energy in the form of normal vibration modes. To observe such energy, if one were to add 2 flat conducting plates very close together, the plates would act like boundary conditions limiting the available vibrations modes; thus, the zero-point energy between the plates would be lower than outside the plates causing a small inward force. Such a force has been postulated as the following (Casimir, 1948),

$$\frac{F}{A} = \frac{\hbar c \pi^2}{240 d^4} = 1.30 \times 10^{-18}\,\text{dyn}\,\text{cm}^2\,d^{-4}, \tag{1.25}$$

where d is the distance between the 2 plates and $\hbar = 1.0545718 \times 10^{-34}\,\text{m}^2\,\text{kg}\,\text{s}^{-1}$ is the reduced Plancks constant. This theoretical result was later verified (Sparnaay, 1957) by measurements of the force per area being $2.50 \pm 1.5 \times 10^{-18}\,\text{dyn}\,\text{cm}^2\,d^{-4}$. With that said, The Casimir Effect being evidence of zero-point energy has been criticised (Bordag et al., 2001) as one can get to the same theoretical prediction through the use of the van der Waals forces (Jaffe, 2005). While this is the case, the simplest explanation is the zero-point one. Thus the Casimir Effect continues to hold under Occam's razor.

There would be many other theoretical contributions towards this vacuum energy, such as the potential energies of scalar fields such as the Higgs Field (Bass and Krzysiak, 2020) but the most dominant would be the zero-point energy which is up to the Plank scale (the scale of which Quantum Field Theory works up to) one predicts a quantum fluctuations energy of $\rho_\Lambda^{PL} \sim (10^{18}\,\text{GeV})^4$. However, when one observes the vacuum energy density required to explain the accelerated expansion, we observe we produce $\rho_\Lambda^{\text{obs}} \leq (10^{-12}\,\text{GeV})^4$ (Carroll, 2000) This means the theoretical prediction is 120 magnitudes larger than observational limits (Weinberg, 1989). This is one of the most significant errors in all physics (although this discrepancy has been reduced from 120 magnitudes to the still very high 30 if one expresses the energy density as a Mass $\rho_\Lambda = M_{\text{vac}}^4$).

This issue has led to many alternative theories being studied to eliminate the constant. For example, generalised '$f(R)$' General Relativity (Capozziello et al., 2018) looks for modifications to the field equations themselves. Quintessence (Capozziello et al., 2003) which adds a single slow-rolling scalar field with minimal coupling to gravity, producing a potential that acts similarly to the cosmological constant and the inclusion of string theory and attempts at a unifying theory for quantum gravity (Polchinski, 2006),(Dvali et al., 2000). Nevertheless, none have yet produced convincing results or made observable predictions.

1.4 Standard model

Through the Friedmann equations and the cosmological constant, we now have a framework in which a homogeneous and isotropic universe can not only expand but accelerate in its expansion. However, they are not yet solvable in this state as we have 3 unknown variables P, a, ρ. The pressure, scale factor and density, respectively. To solve 3 variables, a 3rd equation is required, which comes in the form of the continuity equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0, \tag{1.26}$$

which can be derived directly from the 1st law of thermodynamics $dE = -PdV + dQ$ or by applying the energy conservation equation $\nabla_\nu T^{\mu\nu} = 0$ to the fields equations. However one still cannot solve 3 variables as not all 3 equations are linearly independent (this is proved by the fact that one can combine equations 1.20 and 1.21 to produce equation 1.26). Instead, one has to use a property of a perfect fluid that has a barotropic equation of state. The pressure can be described as a function of density exclusively,

$$P = w\rho c^2, \tag{1.27}$$

where w is a dimensionless constant. Although more commonly known as $P = w\rho$ when $c = 1$ By substituting P with equation 1.27 in equation 1.26 and integrating one produces a (ρ, a) relationship,

$$\rho_i = \rho_{i0}\left(\frac{a}{a_0}\right)^{-3(1+w_i)}, \tag{1.28}$$

where $\rho_i 0$ and a_0 are the present values of the density and scale factor, respectively (though usually $a_0 = 1$), and i represents the different i-fluids. The equation refeq: relationship is indeed a solution to the differential Friedmann equations, but they are useless without understanding the different density ρ_i components and their respective equation of state constants w_i.

1.4.1 Densities

Different components of the universe will have different equations of state. The most obvious is that we see radiation and matter within our stars. To define their equations of state, we assume matter acts as an ideal gas. An ideal gas assumes point particles (they take up no volume) randomly moving with little inter-particle interaction except for direct contact. Both are acceptable for cosmological use as the vast volume of space means any matter is effectively point-like. The same can be said for randomly moving in any direction as the

universe is isotropic (no preferred direction), and little interaction as the space between matter is so significant that any gravitational interaction would be minimum. It can be considered a direct collision when it would not be minimum.

The ideal gas law states that $PV = nRT = NkT$ where P,V,n,R,T,N,k are pressure, volume, number of moles, the gas constant $= 8.31\,\mathrm{J\,mol^{-1}\,K^{-1}}$, temperature, number of particles and the Boltzmann constant $= 1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ respectively. the law can be rearranged in terms of pressure and density ρ,

$$P = \frac{\rho}{\mu} kT, \tag{1.29}$$

where μ is the average particle mass. By using Newton's second law $F = ma$ where F,m,a are force, mass and acceleration one can define $PV = N\mu\bar{v}^2/3$ where $\bar{v}$ is average particle velocity. Combined with the ideal gas law leads to $kT = \mu\bar{v}^2/3$, which gives a final particle pressure of

$$P = \frac{1}{3}\rho\bar{v}^2. \tag{1.30}$$

Relating the equation 1.30 to the equation of state 1.27 gives a particle equation of state-dependent on ρ and $\bar{v}$

$$w = \frac{\bar{v}^2}{3c^2}. \tag{1.31}$$

We categorise the density by the assumed $\bar{v}$. In the case of matter, we state that the velocities are non-relativistic ($\bar{v} << c$) whereas radiation velocities are relativistic ($\bar{v} \approx c$), leading to the following equations of state:

$$w = \begin{cases} 0, & \text{matter} \\ \frac{1}{3}, & \text{radiation} \end{cases}. \tag{1.32}$$

As stated previously, the cosmological constant can be interpreted as a density $\rho_\Lambda = \Lambda/8\pi G$ and thus, under our current model, must also have an equation of state. Due to Λ being constant (no time dependence) one can differentiate its density by time ($\dot{\rho} = 0$) which when applied to the continuity equation 1.26 reveals that $\rho = -P$ and thus

$$w_\Lambda = -1. \tag{1.33}$$

To keep consistent with how we have interpreted the components above, one can also write the universe's curvature in terms of density. In the same vein, as previously shown, we can

rearrange the equation 1.20 to give the curvature energy density ρ_k,

$$\rho_k = \frac{3k}{8G\pi a^2}. \tag{1.34}$$

This shows that $\rho_k \propto a^{-2}$ which when related to our solution from the equation 1.28 implies an equation of state of $w_k = -1/3$.

In terms of usable parameters within a ΛCDM model one usually does not use the raw density ρ instead it is usually normalised to the critical energy density $\rho_{cr,0} = 3H_0^2/8\pi G$ which is the density value one gets from equation 1.20 if we set $k = 0$ and only account for present-day ($a = a_0 = 1$) and include ρ_Λ within ρ. Current values based on our understanding of H_0 have this value at $\rho_{cr,0} = 1.878\,h^2 \times 10^{-29}\,\mathrm{g\,cm^{-3}}$.This allows us to define our density parameter as

$$\Omega = \frac{\rho}{\rho_{cr,0}}. \tag{1.35}$$

Under this definition, all the density parameters (matter, radiation, constant and curvature) should add up to 1, provided we live within a flat universe. We can now use the Friedmann equations with our equations of state, density solutions from equation 1.28 and the redshift-scale factor relation from equation 1.7 to understand how energy densities have changed with expansion and redshift,

$$H = H_0\sqrt{\Omega_{r0}(1+z)^4 + \Omega_{m0}(1+z)^3 + \Omega_{k0}(1+z)^2 + \Omega_\Lambda}. \tag{1.36}$$

Constraining the values of these present-day density parameters is a large part of current cosmology. Especially in the case of weak lensing (see Chapter 2), Ω_m is of utmost importance, so it is worth clarifying. Ω_m is the parameter for what is usually known as 'cold matter' or 'dust' where the cold in its name comes from being of low-kinetic energy compared to its mass-energy (or, as we saw earlier, has very low non-relativistic velocities compared to the speed of light. Matter itself can also be split up into 2 sub-components $\Omega_m = \Omega_b + \Omega_{\mathrm{CDM}}$ where b and CDM refer to baryons and cold dark matter respectively.

Baryonic matter refers to everything in the universe one might classically consider as 'stuff', from every star to every planet, although most of it exists as diffused hot gas around galaxies. In cosmology, we consider baryons to be all the neutrons, protons and electrons, although strictly speaking, electrons are leptons in the context of particle physics. The guiding principle to what is considered a baryon is that their velocities are non-relativistic (so do not include neutrinos), interact with light, and can be directly observed.

On the other hand, what we cannot directly observe through light interactions is Cold Dark Matter Ω_{CDM}. Very little is known about CDM besides that it interacts gravitationally but not electromagnetically. The reason for its introduction (and its importance going forward in this thesis) is that current observations of the universe cannot be explained through the gravitational attractions of baryonic matter alone. The most obvious is the rotation curves of galaxies. This refers to the orbital velocity of stars around a galaxy. To mathematically simplify, if we assume that Newtonian gravity is enough to explain the motion of stars within and around a galaxy, most of the galaxy's mass is concentrated within a laboratory radius R. Under these assumptions, one would expect the orbital velocity outside of this region to be the following:

$$v_{\text{cir}}^2 = \frac{GM}{r}, \tag{1.37}$$

where v_{cir} is the orbital velocity, G is Newton's Constant, M the galaxy's mass and r the radius from that galaxy. One can see that the $v_{\text{cir}} \propto 1/\sqrt{r}$. This does not, however, match with observations. As we can see in Figure 1.2, where the bold line is the rotational curves of NGC (New General Catalogue) 6503, a dwarf spiral galaxy approximately 17 million light-years away (Young, 2017) v_c does not go down at $r^{-1/2}$ but instead remains constant. If one were to relax our assumption of all mass being within radius R (the disk) and allow gas on the outskirts, the gas contribution is still insufficient. Instead, one must model a halo (dot-dash) of 'extra mass' (Dark Matter) to explain the curves. This observation is not limited to HGC 6503 but has been observed in 1000s of galaxies (Catinella et al., 2006) including our own, the Milky Way (Sofue and Rubin, 2000). Tracing the 21cm line of atomic hydrogen has allowed us to observe the constant nature of the velocity curves to continue for at least 10s of kpc away from the galaxy (Munoz et al., 2006). This evidence indicates that almost all galaxies contain at least 4 times as much dark matter as they do stars, gas, and dust (Gelmini, 2015) and in the case of smaller dwarf or low surface brightness galaxies, as much as 95% of their mass is DM (Young, 2017).

One of the strongest direct observations of dark matter allows us to relax our Newtonian gravitational assumptions. This direct observation was the 'bullet cluster' (Clowe et al., 2006). The cluster 1E-0657-558 (Tucker et al., 1998) contains 2 main galaxy concentrations that have recently collided ($\sim 100\,\text{Myr}$ ago). We presume they have recently collided as their X-ray-emitting baryonic gas is concentrated in the cluster's centre, whereas their visible light-emitting galaxies are still separated. This is consistent with a model in which galaxies act like collisionless particles (as the chance of a direct galaxy collision is so low), whereas the gasses around the galaxies experience ram pressure slowing them down. Under these conditions, the galaxies have been decoupled from the X-ray plasma. Thus, observations of the cluster's gravitational potential should be at the centre of the cluster where the plasma

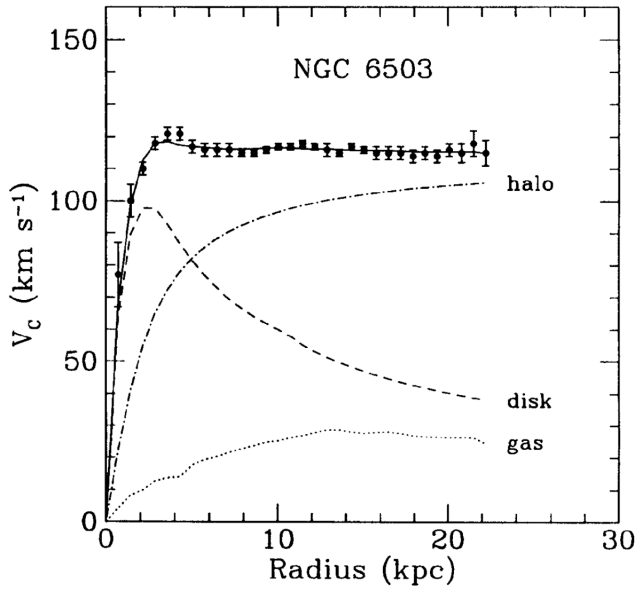


Figure 1.2: In rotational curves of galaxy NGC 6503, the dark line is observed while the dashed lines are the theoretical components (Freese, 2008).

(and thus the majority of the cluster baryonic mass) exists. An observation of the gravitational potential can be made through weak lensing (see chapter 2), which gives us a 2-D κ-map of the cluster as seen in Figure 1.3. The κ-map roughly follows a map of the gravitational potential; thus, one can see the gravitational potential is placed with the galaxies, not the plasma. The spatial offset between the centres of the baryonic mass peaks and the centres of the total mass is detected to 8σ significance.This contradicts a baryonic-only universe and indicates that most of a cluster's mass is with DM halos surrounding the galaxies. As with the rotation curves above, this is not an isolated event. In fact, through analysing data from Chandra and Hubble Space telescopes, an investigation took place on 72 cases of galaxy cluster collisions (Harvey et al., 2015) all at different angles and times about earth observers. This is crucial as a criticism of interpreting a single observation is that due to Figure 1.3 only being a 2-D image, the effects may be caused by unseen masses along the line of sight. Due to the 72 case analysis, they concluded to 7.6σ significance that most of the cluster's mass lies around the galaxies as halos as opposed to the surrounding plasma. The cluster merger studies not only provide compelling evidence for the existence of DM but also provide an insight into its physical nature. As the DM from the 2 galaxy concentrations passed through each other with no observable drag, it indicates that DM not only does not interact with matter (as it passed through the gas unaffected), but it also does not self-interact (as it passed through more DM). If one was to interest DM as a particle, one can measure

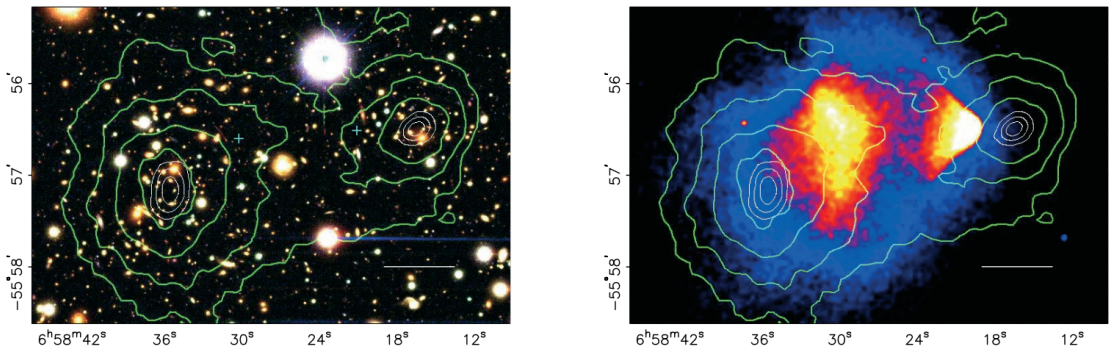


Figure 1.3: *left*: Optical image of merging cluster IE 0657-558, the white bar indicates 200kpc while the blue plus the centre of the plasma clouds, the green contours are the κ maps, and the white contours are errors in the κ-peaks at 68.3% 95.5% and 99.7%. *right*: Same as left but with a 500ks Chandra image indicating the position of the plasma. (Clowe et al., 2006).

its self-interaction cross-section based on these 72 cases to be $\sigma_{DM}/m = 0.47\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ to 2σ significance. Above were 2 of the most significant findings in evidence for DM, but others do exist as well, such as the observation of hot gas in the centre of galaxy clusters such as the Coma Cluster (Briel and Henry, 1997) were without the presence of DM creating potential well the hot gas would have evaporated out the cluster. Observations of the radial velocity dispersions of galaxies also in the Coma Cluster showed them not to be explainable with the visible matter alone (Zwicky, 1937).

Although evidence supports the existence of DM, the fact that it is so weakly interacting (both with itself and matter and photons outside of gravitational) and that it has not been detected within a laboratory means our understating of what DM is considerably low. Thus a level of scepticism over its existence still exists. Furthermore, there are some challenges to our current understanding of DM. Areas where observation contradicts our current DM model (Weinberg et al., 2013). A case in point is the 'core-cusp' problem; current cosmological simulations predict that DM halos have a 'cusp-like' structure where the dark matter distribution is sharply peaked at the centre. In contrast, observations of galactic velocity rotation curves indicate a flat constant core dark matter profile. Thus, our understanding of DM predicts too much matter in the 1st few central kpcs of typical galaxies (Alam et al., 2001). Another issue is most well known as the 'missing satellite' problem. CDM is predicted to preserve fluctuations even at small scales due to its non-relativistic and weakly interacting nature. These preserved small-scale fluctuations should be observable as satellite galaxies orbiting large galaxies (such as the Milky Way) of the order of 50-200 (Moore et al., 1999). Pre-2000 current observations contradicted this prediction, with only 9 faint dwarf galaxies being detected to be orbiting the Milky Way (Klypin et al., 1999). However, in the last 2

decades, further observations by SDSS (Willman et al., 2005) and DES (Nadler et al., 2019) have found enough satellite galaxies in small patches that one can extrapolate that across the total halo volume, the missing satellites would be found (Koposov et al., 2007). Very much related is the 'too big to fail' problem (Boylan-Kolchin et al., 2011a) which again looks at satellite galaxies and sub-halos. In the same way that CDM would preserve small fluctuations, it would also preserve slightly larger ones. Thus one expects to find massive sub-halos. However, current observations of the Milky Way have found that the largest sub-halos we are aware of are still 5 times smaller than expected (Boylan-Kolchin et al., 2011b). This problem is not restricted to the Milky-Way either, as similar discrepancies have been found in our closest galactic neighbour, the Andromeda galaxy (Tollerud et al., 2011). Although the argument can be made that we have not observed the massive subhalos because they are 'dark,' i.e. do not contain bright dwarf galaxies, this goes against our understanding that massive halos host galaxies seem unlikely.

Due to the problems mentioned above and the continued issue of a lack of CDM particle confirmation, alternative ideas to CDM have been theorised. Modified Newtonian dynamics (MOND) (Milgrom, 1983) is the most notable. The reasoning behind MOND is not to remove CDM but to explain why the gravitational fields act as if they require more mass than the baryons offered by the baryons alone (Famaey and McGaugh, 2011). The idea behind the modification is that our understanding of gravitational dynamics is limited to the high acceleration environments of our solar system and that outside the core of our galaxy, stars accelerate at a much slower pace; thus, Newton's 2nd law of motion can be rewritten:

$$F_g = \begin{cases} \frac{GMm}{r^2}, & a \gg a_0 \\ \frac{GMma_0}{ar^2}, & a \ll a_0 \end{cases}, \tag{1.38}$$

where F_g, M, m, r, a is the gravitational force, the masses of the 2 objects, and their relative radius and acceleration of 1, whereas $a_0 \sim 10^{-10}\,\mathrm{m\,s^{-2}}$ is the typical acceleration of stars in a galaxy. In other words, in the regime of the solar system, $a \gg a_0$ Newtonian dynamics stay touched, whereas, for galactic dynamics (especially those on the outskirts), the dynamics change, leading to a new centripetal velocity of

$$v_c(a \ll a_0) = (a_0 GM)^{1/4}, \tag{1.39}$$

a velocity independent of r can lead to the constant nature of observed velocity curves.

Relativistic extensions to MOND have also been suggested, such as the Tensor-vector-scalar theory (Teves) (Bekenstein, 2004) which attempts to make the theory compatible with structure formation. The modifications would solve issues such as the velocity curves without

needing CDM. However, the theory has been tightly constrained due to observable evidence (the Bullet Cluster), which successfully decouples the mass of the galaxies from its baryonic matter, which is incompatible with MOND. Many MOND alternatives would still require their CDM equivalent to agree with the most recent observations (Ferreras et al., 2012).

Given these discrepancies, we will continue with the primary interpretation of DM, an unknown particle. However, given our current understanding, we can make some assumptions about the type of particle we expect DM to be:

- It would be a particle that does not carry electrical or colour charges (due to lack of photon interaction evidence and known exotic isotopes).
- It would have been produced early in the universe (as required with current simulations). Thirdly would have to be stable to be here still today.
- They would interact gravitational (to solve all the issues seen above) but self-interact incredibly weakly (bullet cluster).

The candidates for the dark matter particle (sometimes referred to as χ) can be roughly classified into camps. WIMPS (Weakly Interacting Massive Particles) and non-WIMPS. WIMPS usually have masses in the range of $m_{\text{WIMP}} \sim 10\,\text{Gev} - \text{Tev}$ and are so far the studied group due to them appearing in other particle theories, having the correct relic density and offering the most straightforward chance of confirmation by the detection (Feng, 2010). Below are the 2 leading WIMP-type candidates :

- Neutralinos. A popular particle physics theory beyond the standard model is supersymmetry (Martin, 1997) which introduces a new set of particles that only differ in their spin from our current particles. The neutral spin 1/2 new particles can mix to form other particles, one of them has a mass and interaction cross-section inline current WIMP models
- Kaluza-Klein Dark Matter. Particles produced through the universe having more spatial dimensions in a theory known as Universal Extra Dimensions (USD) (Appelquist et al., 2000). One of these extra particles is light and stable enough to be a WIMP candidate.

Others exist, such as T-odd particles (Birkedal et al., 2006) and Branons (Cembranos et al., 2003), but much less work has focused on them. Below is a summary of non-WIMP candidates that could also be (or contribute towards) dark matter.

- Axion. Originally proposed to solve the strong CP problem, a new pseudoscalar field is introduced leading to an extremely light and weakly interacting particle (Asztalos et al., 2006).

- Sterile Neutrinos. One can introduce right-handed neutrinos for their mass terms to be allowed. They must not interact with any standard model particles (thus sterile).
- Weak-Scale Gravitinos. An example of a SUPERWIMP (Feng et al., 2003). Like WIMPS, they have the correct relic density but much weaker interactions. They exist within supersymmetry as the spin 3/2 partner to the Graviton (Ellis et al., 2003).

As with WIMPs, countless more candidates exist, especially in the realm of 'hidden dark matter' (Feng, 2010) where particles can exist without standard model interactions.

It is worth noting that outside of new gravitational theories and individual unknown particles, there is a third possibility of explanation for the matter discrepancy in the form of large astronomical objects. Examples of said objects include Massive compact halo objects (MACHOs), which suggest that dark matter may be very dim stars, such as brown dwarfs (De Paolis et al., 1998) that have so far been undetected. However, recent microlensing surveys have ruled out the possibility that MACHOS can substantially contribute to dark matter (The MACHO collaboration et al., 2000). Similarly, Massively Compact Objects, which are dense pieces of non-radiating matter, such matter should be detectable in the lensing of supernovas, but current supernova observations have ruled out their possible significance (Metcalf and Silk, 2007). Lastly, primordial black holes produced in the early universe are small enough not to be ruled out as DM candidates. These possibilities for non-particle explanations have already been mostly ruled out or considered unlikely so we will continue with the general particle explanation.

Regarding finding these possible Dark Matter Candidates, the experiments can be loosely classified into 3 categories; direction detection, indirect detection and laboratory production (Heros, 2020). Therefore, we will briefly explore these 3 classifications.

Direct detection experiments use presumed dark-matter-nucleus interactions to produce nuclear recoils. The number of recoils one gets (does not get) helps constrain possible dark matter particle mass and cross-section (m_χ,σ_χ). The most known example is DAMA/LIBRA (Bernabei et al., 2018) a detector 1400 m underground in Italy. This detector takes advantage of the earth's annual modulation in velocity around the milky way due to it orbiting the sun (as orbital velocity and the velocity of the solar system are not always in the same direction. Thus the number of dark matter interactions on the earth will also modulate annually. A modulation that has now been detected to 9.5σ. These modulations, dark matter detections, have been under scrutiny because they suggest a WIMP mass and cross-section that has so far been disfavoured by every other experiment (Heros, 2020). This can be seen in Figure 1.4 where the DAMA results (pink contours) fall in disfavoured regions of mass and cross-section (green). Such contradictions have been met with opposing arguments. One of them suggests the modulations in the DAMA data are caused by cosmic ray background (Blum, 2011) not

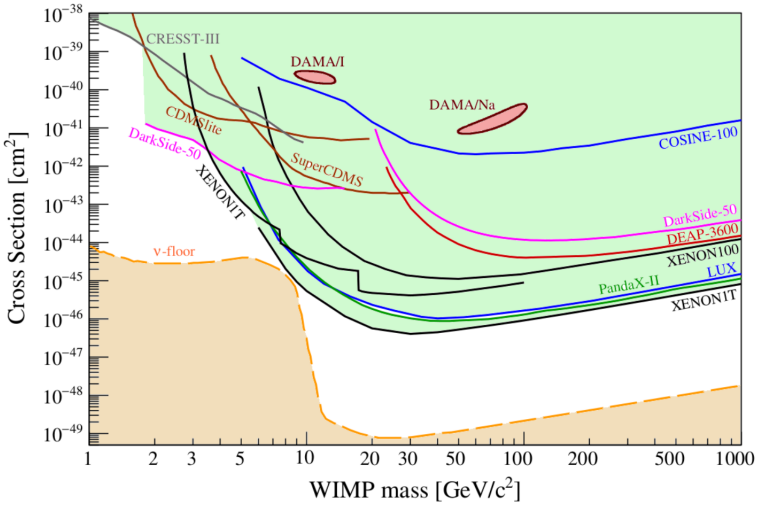


Figure 1.4: Experimental limits on current Dark matter - nucleus detection experiments. Regions above lines are disfavoured at 90% confidence. The regions labelled 'DAMA' are the detections of the DAMA/LIBRA experiments (Heros, 2020).

Dark matter, and another suggests that the DAMA experiment's unique choice of Isotopes that contain an odd number of protons, i.e. Na(11) and I(53) are the reason they are detecting Dark Matter. In contrast, similar even number proton experiments do not (Arina et al., 2015).

Other examples of direct detection include Cryogenic Dark Matter Search (CDMS(Agnese et al., 2013), the Germanium Neutrino Technology (CoGeNT) (Aalseth et al., 2013) and the Cryogenic Rare Event Search with Su- superconduction Thermometers (CRESST) (Angloher et al., 2012). All of which to date have ether constrained m_χ,σ_χ further or had detected ions with extremely conflicting results. However, new detectors such as the European Underground Rare Event Calorimeter Array (EURECA) hope to explore lower mass regions, yet unchecked (Kraus et al., 2009).

Alternatively, one can use indirect detections. These detections presume that as Dark matter gravitation attracts more dark matter in heavy objects (galaxies and stars etc.), the Dark matter will either annihilate itself (as Dark matter may well be its anti-particle) or decay. Either way, this should lead to a detectable flux of photons, neutrinos or cosmic rays (Kunnen, 2015). Using the centre of the sun (Murase and Shoemaker, 2016) or earth (Aartsen et al., 2013) as sources of this neutrinos flux are desirable for detectors such as neutrino telescopes for 2 reasons. Only neutrinos would escape the centres, and neither the sun nor the earth produces high-energy neutrinos, thus reducing background noise. In both cases one has to assume that the mass has reached an equilibrium between capturing exterior dark matter and annihilating interior dark matter. As seen in previous studies, this assumption limits the dark matter mass and cross-section. Figure 1.5 shows such limits from 3 such detectors as

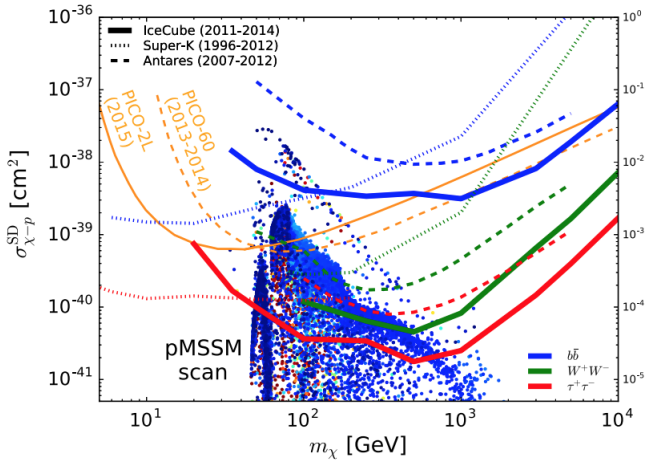


Figure 1.5: Experimental limits on current Dark matter - proton cross-sections from indirect detection (Aartsen et al., 2017).

ANTARES (ANTARES collaboration et al., 2016), Ice-Cube (The IceCube Collaboration et al., 2015) and Super-K (Choi et al., 2015) for the full annihilation to $b\bar{b}, W^+W^-, \tau^+\tau^-$.

For the case of earth indirect detections, one has to assume a value for the annihilation cross-section to constrain the dark matter-nucleon cross-section. Alternatively, one can search for dark matter through galactic centres where one searches for increased fluxes in gamma rays and neutrinos, such as with the High Energy Stereoscopic System (HESS), which uses five Imaging Atmospheric Cherenkov Telescopes located in Namibia to observe high-energy gamma-rays from the galactic centre of the milky way (Moulin, 2019). Cosmic ray detections have also been made that can be attributed to dark-matter annihilation, but more straightforward non-dark matter explanations involving supernovae remnants have also been suggested.

As stated, indirect detection of gamma rays, cosmic rays and neutrinos from the centre of galaxies brings about an issue not found with an indirect observation from the sun or centre of the earth; background noise. This is mainly due to uncertainties in the many astrophysical processes between the galaxy's centre and our observation. This issue may be solved through the search for anti-deuterium $\bar{d}$, a secondary particle produced by WIMP annihilation associated with minimal amounts of background. This is mainly because the calculated peak production of $\bar{d}$ by dark matter takes place at low kinetic energies, whereas the background $\bar{d}$ produced by cosmic rays are at much higher energies. Such a difference in kinetic energy makes deciphering, which is caused by dark matter, much more accessible, thus significantly reducing background noise (Donato et al., 2000). Due to this lack of background noise, the search for $\bar{d}$ is being considered as a viable next step in the constraining of dark matter particle parameters (Randall and Xu, 2020) and thus has many current efforts trying to detect

$\bar{d}$ such as the General Antiparticle Spectrometer (GAPS) (Aramaki et al., 2015) and the AMS-02 (Cholis et al., 2020).

The final form of detection is from laboratory production or, more specifically, from collider experiments. Large particle colliders constantly search for new particles that could include new dark matter candidates or those already mentioned above. Proposals are currently being considered at multiple sites, including The European Organization for Nuclear Research (CERN), Jefferson Lab (JLAB), Fermi National Accelerator Laboratory (FNAL) and the Stanford Linear Accelerator Center (SLAC) (Heros, 2020). Such particle detectors would offer a level of precision simply not available from the other methods of detection mentioned above. However, there are some significant limitations in this classification. Firstly since the indicator of a dark matter interaction is a change in expected energy, many model assumptions have to be made to connect the energy change to a dark matter cross-section (as opposed to direct detection, which measures the cross-section directly). Secondly (and potentially more damning), any discovery made through collider experiments would discover, at best, a dark matter candidate but would not be able to link that candidate to the actual substance currently holding our galaxies together without coupling it with an astrophysical observation. With this in mind, direct and indirect observation are still the best case for discovering dark matter going into the future.

1.4.2 Distance Measures

As redshifts are the main focus of this thesis, it is worth noting that they are not an allegory for distance but a measure of how much expansion has occurred between the photon being emitted and observed. The exception is at low redshift when one can produce a linear relationship between redshift and distance

$$z \approx \frac{d}{D_H}, \tag{1.40}$$

where d is distance and D_H is the Hubble distance which is defined as

$$D_H \equiv \frac{c}{H_0} = 3000h^{-1}\text{Mpc} = 9.26 \times 10^{25} h^{-1}\text{m}, \tag{1.41}$$

where h is a dimensionless parameter that expresses our uncertainty in H_0 thus $(0.6 < h < 0.8)$. However, usage of this distance approximation may not apply in cosmological studies, especially at high redshift (Fairall et al., 1992). The relationship between redshift and expansion becomes very parameter-dependent. This becomes clear as we redefine the

equation1.36,

$$E(z) = \sqrt{\Omega_{r0}(1+z)^4 + \Omega_{m0}(1+z)^3 + \Omega_{k0}(1+z)^2 + \Omega_\Lambda}, \tag{1.42}$$

where $H(z) = H_0 E(z)$. This defines a function dependent on redshift that's proportional to the logarithm of the scale factor. As this expansion rate changes with time and redshifts, the definition of 'distance' is becoming less defined. Depending on the method used to measure distance, one may acquire different values entirely at different points in time, as shown below.

The closest one can probably get to an accurate distance measure is comoving distance, which is the distance between 2 objects if we ignore the main complication, which is the scale factor. Suppose we consider the simplified case, a galaxy in our line of sight of known redshift z. Then, we can redefine our equation 1.5 in the comoving coordinate χ,

$$ds^2 = -cdt^2 + a^2 d\chi^2. \tag{1.43}$$

Taking the case of light, where $ds = 0$ (a null geodesics), we can formally define the comoving distance between 2 radial objects,

$$D_C = \chi = \int_{t_{emt}}^{t_0} \frac{cdt}{a} = \int_a^1 \frac{cda}{a^2 H} = \int_0^z \frac{cdz}{H} = D_H \int_0^z \frac{dz}{E(z)}, \tag{1.44}$$

where $dt = da/aH$ and $da = -a^2 dz$. It is clear that when we consider low redshift distances where $E(z) \sim 1$, then the comoving distance rediscovers the linear relationship defined in equation 1.40.

We can now consider different types of measures. One such example is the angular distance. If we consider the object in the sky of known diameter by measuring the angle of separation across its diameter, we can work out how far away it is as simple as $d = \text{diameter}/\theta$ given $\theta \ll 1$. We can then relate the diameter distance (angular distance) d_A to the comoving distance

$$d_A = a(t)\chi = \frac{D_C}{1+z}. \tag{1.45}$$

This particular equation shows the looseness of which something can be defined as the 'distance' as from Figure 1.6.a one can see that dependent on your model, objects of known sizes that appear to have the same diameter when observed can be at 2 completely different redshifts. In fact, after a point, objects that appear angular distance-wise 'closer' have larger redshifts.

Similar complications appear for another measure of distance; the luminosity distance. This distance refers to the idea that objects further away appear less bright; thus, if an object's

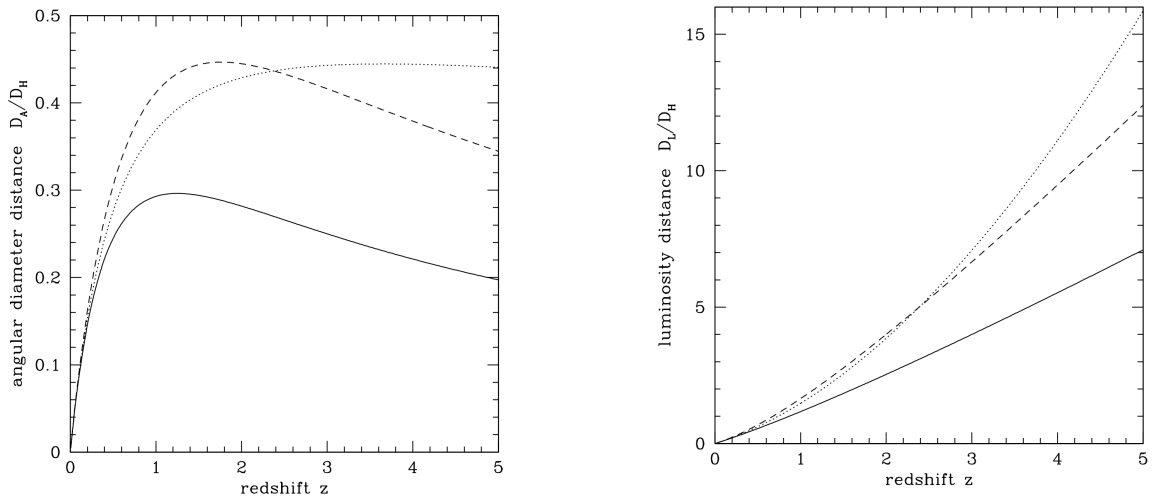


Figure 1.6: The ratio of different distance measures to the Hubble distance against redshift. The curves represent different models of $(\Omega_M, \Omega_\Lambda)$, solid (1.0,0.0), dotted (0.05,0.0), dashed (0.2,0.8). *left*: angular diameter distance. *right*: Luminosity distance (Hogg, 1999).

intrinsic brightness is known and compares it to the observed brightness, one can deduce its distance (just as the angular distance uses a known intrinsic size). More precisely, we define the luminosity distance D_L as the distance that follows the following relation between the object has observed flux F and its intrinsic luminosity L:

$$F_{obs} = \frac{L_{\text{int}}}{4\pi D_L^2}. \tag{1.46}$$

Under the assumption that the photons spread equally along the surface of a sphere around the object. Due to the expansion of the universe, the luminosity of the object does stay constant, with the intrinsic luminosity defined as $L_{\text{int}} = dE/dt$ and the observed luminosity as $L_{\text{obs}} = dE_0/dt_0$ 2 factors must be considered. Firstly as stated in equation 1.7 photons redshift so their observed energy decreases as $dE/dE_0 = 1/1+z$. Secondly, time intervals also change depending on expansion, as shown with the equation 1.43when $ds = 0$ if the universe expands, so are the time intervals as $dt/dt_0 = (1+z)$ and so one can relate both luminosities as a function of redshift:

$$L_{\text{obs}} = \frac{dE_0}{dt_0} = \frac{dE}{dt(1+z)^2} = \frac{L_{\text{int}}}{(1+z)^2}. \tag{1.47}$$

Presuming a photon has travelled with a comoving distance D_c accounting for the extra factors, this would give the following flux relation

$$F = \frac{L_{\text{obs}}}{4\pi D_c^2} = \frac{L_{\text{int}}}{4\pi(D_c(1+z))^2}, \tag{1.48}$$

where we have omitted the modern-day scale factor as we presume $a_0 = 1$. By relating to the equation 1.46, we can define the luminosity distance as a function of redshift

$$D_L = D_c(1+z). \tag{1.49}$$

As with angular distance, the redshift dependency means the change in flux one expects changes with the cosmological model as seen in 1.6.b. The angular and luminosity distance can be related through Etherington's distance duality relation (DDR) (Etherington, 1933)

$$D_L = (1+z)^2 D_A. \tag{1.50}$$

DDR represents a testing ground for some of the cosmologies' more fundamental assumptions. Due to DDR being a ratio only dependent on redshift, it should hold regardless of choices of $\Omega_{m/\Lambda/\gamma/\kappa}$. The relation only requires photon numbers to be conserved, photons to follow null geodesics and space-time to follow a metric. An expanding universe is not even required as this would lead to $z = 0$ for all areas of space, making $D_L = D_A$; thus, any observations against the relation would be significant.

Various forms of parameterisation have been suggested to investigate the validity of DDR. Previously DDR has been parametrised using $D_L = D_A(1+z)^{2+\varepsilon}$ where DDR would be valid for the case of $\varepsilon = 0$. Such an example of this approach was through the testing of cosmic-opacity (Avgoustidis et al., 2010); DDR relies on the conservation of photon counts as a change in photon flux would not change angular distance but would change luminosity so one can determine that the observed luminosity distance may be affected by the photons being lost from source to observation

$$D_{L,\text{obs}}^2 = D_{L_{\text{true}}}^2 e^{\tau(z)}, \tag{1.51}$$

where $\tau(z)$ is an opacity parameter that to first-order approximation is $\tau(z) = 2\varepsilon z$. Such change in flux can be due to obvious astrophysical sources such as interactions with dust. However, the clustering nature of said sources can be mostly avoided, leading to a sub per cent level of constraint on their effects. Instead, the study of cosmic opacity comes from the possibility of exotic particles coupling with photons in a so-far unknown way. D_L measurements of SNe type 1a were compared to measurements of $H(z)$ so that in the

redshift range of $0.2 < z < 0.35$ cosmic opacity was constrained to $\tau < 0.012$ which leads to a DDR parameter constraint of $\varepsilon = -0.04^{+0.08}_{-0.07}$ (Avgoustidis et al., 2010) thus expressing no violation of DDR.

The above example was particular. Modern studies use a more generalised function to represent possible violations in DDR, mostly through the use of a $\eta(z)$ function

$$\frac{D_L}{D_A}(1+z)^{-2} = \eta(z), \tag{1.52}$$

where a violation to DDR would be where $\eta(z) \neq 1$. The advantage of such a function is that it opened the doors to trying different parametric models, such as

$$\eta(z) = \begin{cases} 1+\eta_0 \\ 1+\eta_1 z \\ 1+\eta_2 \frac{z}{1+z} \end{cases}, \tag{1.53}$$

where both η_0, η_1 and η_2 must be 0 to ensure DDR.The motivation behind the use of multiple models is that they mimic the possible expression for an equation of state for different dark matter models (Cao and Zhu, 2014) such as one with a constant equation of state (η_0) or more time-varying models (η_1, η_2). The first 2 models are continuous and smooth, whereas the latter has a correction to prevent divergence at high-z. All models must comply with the idea that DDR holds at $z \to 0$. Such examples of work include the study of galaxy clusters, specifically measurements of the Sunyaev-Zeldovich effect (SZE) and X-ray surface brightness to gather DDR values of $\eta_1 = -0.056^{+0.1}_{-0.1}$ and $\eta_2 = -0.088^{+0.14}_{-0.14}$ all in agreement with DDR (Holanda et al., 2011). While most recent studies have suggested a combination of radio quasars and gravitational waves, specifically 120 intermediate-luminosity quasars with redshifts ranging from 0.46 to 2.80 were observed with the Very Large Baseline Interferometry (VLBI) (Cao et al., 2017) to estimate D_A, this was then combined with D_L estimates through simulated gravitational wave observations to find DDR to be verified to a precision of 10^{-2} (Qi et al., 2019). The gravitational waves were simulated as current wave detectors such as the LIGO/Virgo collaboration (Abbott et al., 2016) to detect a limited amount of events to be used for such analysis. Instead, simulations were based on future third-generation detectors such as the Einstein Telescope (ET), which could be making observations as early as 2035 (Maggiore et al., 2019) by which time quasar observations would have expanded to possibly allow precision of $\Delta\eta = 10^{-3}$. The advantage of a gravitational wave-based approach is that one can obtain a luminosity distance that is completely independent

of photon-number non-conservation (Schutz, 1986) allowing constraint on exotic particles alone.

1.4.3 Perturbation

The model, as summarised by the equation 1.36 is enough to make broad statements on cosmological scales. It is even enough to discover the nuances of which different measurements of distance change in value depending on their redshift, as seen above. However, we do not live in an FRW universe mostly because (as stated previously) an FRW universe requires homogeneity. Although this may be true on the largest scales ($\geq 200\,\mathrm{Mpc}$), the fact that we have dark matter haloes, galaxy clusters, and stars is evidence of a disparity in matter densities or, more generally, non-homogeneity. Instead, we model our universe as a *perturbed* FRW universe. This is achieved by presuming the background universe is homogeneous, but, at any particular point in space, they may be a small difference where that small difference is defined within the metric as

$$\delta g_{\mu\nu} = g_{\mu\nu} - \bar{g}_{\mu\nu}, \tag{1.54}$$

where $\delta g_{\mu\nu}, g_{\mu\nu}, \bar{g}_{\mu\nu}$ are the perturbations, true metric and mean homogeneous metric. Such a change in metric would lead to similar changes in the Ricci tensor ie.$R_{\mu\nu} = \bar{R}_{\mu\nu} + \delta R_{\mu\nu}$. Perturbations in the Stress-energy tensor are not as linear as it's components that are perturbed linearly such as its physical density $\delta\rho(\eta,x)$, pressure $\delta P(\eta,x)$ and 4 vectors $\delta\rho^{\mu}(\eta,x)$. Where x is space and η is conformal time coordinate which is defined as

$$\eta = \int_{-\infty}^{t} \frac{dt}{a}, \tag{1.55}$$

so that one can rewrite equation 1.43 in a coordinate system that completely scales with the scale factor.

$$ds^2 = a^2(\eta)[-d\eta^2 + \delta_{ij}dx^i dx^j], \tag{1.56}$$

where δ_{ij} is the Kronecker delta in the case of a flat universe.

For this thesis, the details are not required, mostly because the tensor perturbations are known as gravitational waves, are not coupled to density, and are not important for the large-scale structure we observe in the sky. Instead, we only care for the scalar matter perturbations, which we can redefine as the matter density contrast,

$$\delta = \frac{\delta\rho - \bar{\rho}}{\bar{\rho}}. \tag{1.57}$$

Such a definition allows us to use linear perturbation theory derived from Newtonian dynamics (Ryden, 2006). We consider a perturbation as a sphere of constant density with radius r and mass M slightly denser than the homogeneous background so that $|\delta| \ll 1$. Due to the gravitational attraction of the matter, the system is dynamic with an acceleration at radius R

$$\ddot{R} = -\frac{GM}{R^2}. \tag{1.58}$$

As this is the gravitational acceleration on the surface of the sphere, as long as there is extra mass there ($\delta > 0$), then the sphere will collapse inwards ($\ddot{R} < 0$). Using the fact that the mass can be defined as $M = \pi\rho R^3/3$ (which is required for the conservation of mass), we can substitute ρ for $\bar{\rho}$ in the equation 1.57to obtain the following

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\bar{\rho}(1+\delta), \tag{1.59}$$

as $M(t) \propto R(t)^3\rho$ one can also state that $R(t) \propto \rho(t)^{-1/3} = \bar{\rho}(t)^{-1/3}[1+\delta(t)]^{-1/3}$ we can now consider how the radius changes in the context of an expanding universe by considering the (ρ, a) relationship in equation 1.28 for matter ($w = 0$) to give $\rho \propto a^{-1/3}$.

$$R(t) \propto a[1+\delta(t)]^{-\frac{1}{3}}, \tag{1.60}$$

$$\frac{\ddot{R}}{R} = \frac{\dot{a}}{a} - \frac{1}{3}\ddot{\delta} - \frac{2\dot{a}}{3a}. \tag{1.61}$$

Combining equation 1.59 with equation 1.61 yields

$$\frac{\dot{a}}{a} - \frac{1}{3}\ddot{\delta} - \frac{2\dot{a}}{3a} = -\frac{4\pi G}{3}\bar{\rho} - \frac{4\pi G}{3}\bar{\rho}\delta, \tag{1.62}$$

the equation 1.62 is significant as its the equation of motion of a homogeneous universe with small inhomogeneity. We can separate the homogeneous part from the perturbation part. By setting $\delta = 0$ and thus removing all perturbations we are left with an acceleration equation of pressure-less matter in a homogeneous and isotropic universe $\dot{a}/a = -4\pi G\bar{\rho}/3$ which when removed from equation 1.62 leaves us with the motion of small linear perturbations

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\bar{\rho}\delta = 0, \tag{1.63}$$

this can give us an insight into the effect expansion has on the perturbations in the universe. Consider a perturbed ($\delta \neq 0$) but static ($H = 0$) universe. This would yield an equation of motion of $\ddot{\delta} = 4\pi G\bar{\rho}\delta$ which has an exponential solution. By simply putting small perturbations in a static, pressure-less universe would grow exponentially. This is not the

case in an expanding universe where the $H\dot{H}$ works as a friction term (thus, it is known as 'Hubble friction'), slowing down the perturbation growth.

However, this is a Newtonian approximation at best. The true relativistic equation requires the solving of fluid equations as mentioned above, which leads to the slightly modified equation

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}\Omega_m H^2 \delta = 0 \tag{1.64}$$

By including Ω_m, the equations of motion on the perturbations are now model dependent. We can trace how perturbation may have grown at different stages in the history of the universe by studying equation 1.36. At high redshifts, the universe will be radiation dominated so $\Omega_m \ll 1$ and $H = 1/2t$ this leads to $\ddot{\delta} + \dot{\delta}/t \approx 0$ as its equation of motion has a logarithmic solution (slow growth)

$$\delta(t) = B_1 + B_2 \ln t, \tag{1.65}$$

whereas beyond this time when $z < 0$, the cosmological constant will be dominating so $\Omega_m \ll 1$ and $H = H_\Lambda = constant$ the perturbation equation of motion is $\ddot{\delta} + 2H_\Lambda \dot{\delta} \approx 0$ which has exponentially decaying solution,

$$\delta(t) \approx C_1 + C_2 e^{2H_\lambda t}. \tag{1.66}$$

These equations of motion help us successfully constrain the times when perturbations can grow. Thankfully we can consider a matter-dominated era in the universe evolution where $\Omega_m = 1.$ and $H = 2/3t$ where $\ddot{\delta} + 4\dot{\delta}/3t - 2\delta/3t^2 \approx 0$. For this the solution follows a power law with 2 solutions; a decaying mode t^{-1} and an expanding mode $t^{3/2}$ where perturbation grows with expansion as $\delta(t) \propto t^{3/2} \propto a(t)$

This, of course, has all been in the case where $|\delta| \ll 1$ which can help explain the evolution of density perturbation on an initial mass scale of 10^5 solar masses (1 solar mass $= M_\odot \approx 2 \times 10^{30}$ kg). However, we require much larger perturbations to explain the structure we see in our universe. For example, a cluster of galaxies can have a density contrast of $\delta \geq 100$ in this case where $\delta \gg 1$ studies of growth no longer follow linear perturbation theory, and we are forced to move to the non-linear models where higher orders of the expansion of δ are needed around a homogeneous and isotropic universe. Due to their complexity, we mostly rely on numerical simulations to understand the structure of this regime. However, some analytical solutions do exist. One such solution can give a precise result: the symmetric collapse model (Bryan and Norman, 1998). This model assumes a symmetric density perturbation of radius R with zero velocity at its border. Under these conditions, we can treat the perturbation as its FRW universe and apply the Friedmann equations directly. When calculated, we find that the symmetric perturbation grows to a

maximum size, detaches from the expansion of the background and begins to collapse until reaching an equilibrium state. This equilibrium state is called the virial state, a state that is stable and self-gravitating and follows the conditions of the virial theorem (Clausius, 1870)

$$U = -2K. \tag{1.67}$$

U and K are the system's total potential and kinetic energy. Assuming energy conservation during the equilibrium state's evolution, one can calculate the non-linear perturbation at virialisation to be $\delta_{vir} \approx 178$. Although very simple and originally applied to a closed universe, the result is encouraging as N-body simulations have shown that defining a density contrast of $\delta \sim 100-200$ can be very successful at defining Dark Matter halos (Planelles et al., 2014). However, structures are not perfectly spherical, and any deviation from spherical symmetry will only worsen with time; thus, although the collapse model may be a good approximation at early times, it is entirely unable to predict some of the very non-spherical structures we observe today (Monaco, 2016).

Alternatively, one can use the Zel'dovich approximation (Scoccimarro, 1998). Under this approximation, we presume the initial displacement of dark matter particles is known or calculable and that its dynamics are entirely determined by this initial configuration of fluid elements so that a particle's Euclidean position is only a function of the initial Lagrangian (initial) position q and time t,

$$r(q,t) = a(t)q + F(q,t), \tag{1.68}$$

where $F(q,t)$ is the displacement field which where local gravitational effects can be seen as $F = \nabla\Phi/4\pi G\bar{\rho}$. Where ∇ is the gradient vector defined as $\nabla f = (\partial f/\partial x, \partial f/\partial y, \partial f/\partial z)$ and Φ is the gravitational potential. Such an approximation allows us to approximate the dynamics of a collapsing or expanding perturbation under local gravitational effects giving a predicted evolution of an over density

$$1 + \delta(x,t) = \frac{\rho(x,t)}{\bar{\rho}(t)} = \frac{1}{(1-c\lambda_1)(1-c\lambda_2)(1-c\lambda_3)}, \tag{1.69}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the eigenvalues to the deformation tensor $\partial f_i/\partial x_j$ which is usually aligned with the axis of an ellipsoidal structure. The Zel'dovich approximation has 2 important advantages over the spherical collapse. Firstly, it makes no oversimplification over perturbation geometry (it can now be elliptical instead of spherical) and can remain accurate further into the non-linear regime. These eigenvalues allow for different structures to form that were simply not possible under the spherical collapse model, such as

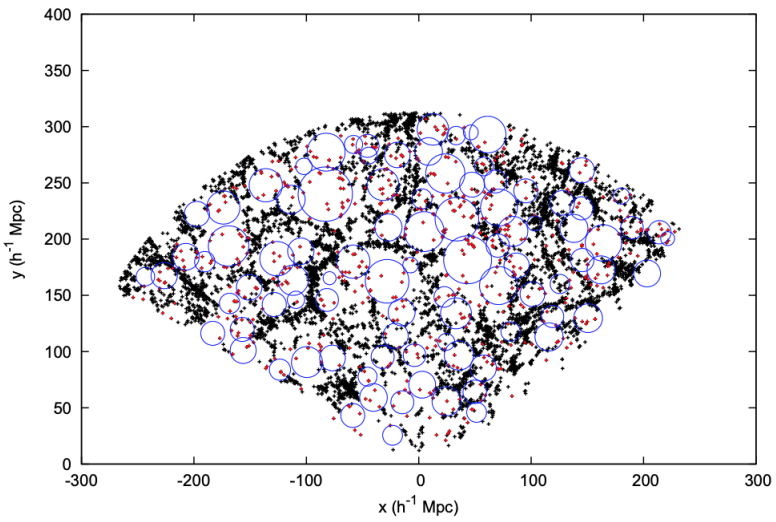


Figure 1.7: Identification of void galaxies (red crosses) within void structures (blue rings) within a projection of a $10h^{-1}$ slab of the SDSS Data Release 9 CMASS sample (Sutter et al., 2014b).

- Void $0 > \lambda_1 \geq \lambda_2 \geq \lambda_3$ expansion along all eigenvectors.
- Sheet $\lambda_1 \geq 0 > \lambda_2 \geq \lambda_3$ otherwise known as a Zel'dovich pancake where the structure collapse along its shortest ellipsoidal axis
- Filament $\lambda_1 \geq \lambda_2 \geq 0 > \lambda_3$ the structure collapses along a second axis from a string-like structure containing a central overdensity
- Halo $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ final collapse along last axis becoming a spherical structure.

Such a structure has not only been predicted by simulation but confirmed by observation. Such an example is the SDSS Data Release 9 CMASS sample(Ahn et al., 2012), a galaxy survey covering 3000 square degrees of sky at redshifts of $0.43 < z < 0.7$ containing 455,281 galaxies. Using a void finding algorithm called VIDE (Sutter et al., 2014a), they could identify 10s of void-like structures filling $\sim 60\%$ of available volume as seen in Figure 1.7. With an effective radius between 20 and $70\,h^{-1}$ Mpc (Sutter et al., 2014b) voids may be among the most significant structures in the universe. Identifying and observing voids opens new doors of possible experimentation, primarily because of their highly low-density environments. Studies of the voids in Figure 1.7 were found to have an average density contrast of $\delta = -0.85$. This allows for general relativity tests in the 'low gravitational' environments of which MOND predicts changes to our current understandings, as well as tests for galaxy formation models seeing how predictions stack up on the low-density side of the spectrum (Peebles, 2001).

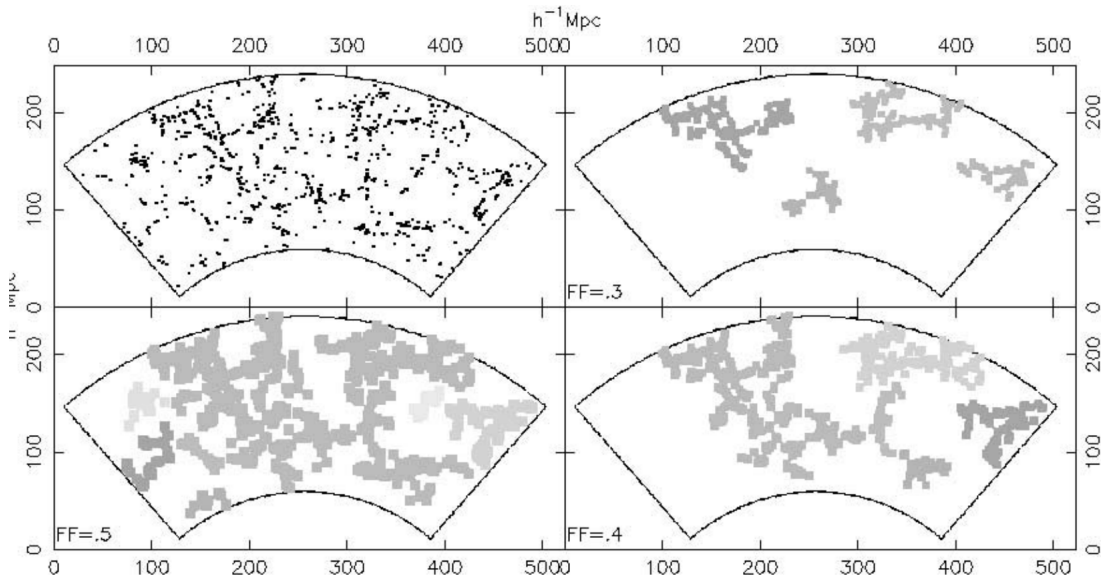


Figure 1.8: *Top Left* : Distribution of uniformly sampled galaxies from LCRS. *Clockwise*: The clusters identified by the FOF algorithm at different FOF radii sizes, at each point, the filaments become interconnected into a 'cosmic web' (Shectman et al., 1996).

Another example of these advanced perturbation structures being observed is in the case of filaments. Filaments are the long string of galaxies that encapsulate the densest regions of our universe, with simulations predicting them to take up only $\sim 10\%$ of the universe's volume but $\sim 40\%$ of all the matter content at $z = 0$. An example of observation would be from the Las Campanas Redshift Survey (LCRS) (Shectman et al., 1996) which recorded the angular positions of redshifts of 26,418 galaxies that were radial up to $600\,h^{-1}$Mpc from us. By taking a subsample of 5073 galaxies and collapsing the survey to a 2-D wedge, a friends-of-friends (FOF) algorithm could be used to identify neighbouring galaxies at varying neighbourhood radii. Through Figure 1.8 one can see that as one increases the FOF radii size, clusters that may be the first perceived as spherical become more string-like in their structure until they begin to interconnect connecting multiple superclusters. The analysis found that the length scale of statistically significant filaments is 70-80 h^{-1} Mpc and thus the most significant structures in the known universe (Bharadwaj et al., 2003). By statistically significant, they meant that although $> 80\,h^{-1}$ Mpc filament structures exist, they arose from chance alignments of projection effects, so they are not considered 'real' objects. Due to their prominence and size, the ability of a cosmological model to produce filaments is an easy test of its usefulness, with current observations of the filamentary structure being in agreement with a flat low Ω_M universe (Sousbie et al., 2008). These tests can be incredibly robust as measurements of the filamentary structure seem to incentivise observation effects such as

redshift space distortions, edge effects and incompleteness (Coil, 2013). While not covered here in as much detail, sheet structures have also been observed through data from the LCRS, revealing rich sheet-like structures surrounding dense regions (thus why void edges seem so well defined) (Doroshkevich et al., 2001). While halo structures seem paramount to any dark matter model of N-body simulation going forward.

1.4.4 Power spectrum

Although non-linear perturbations are prevalent and observable given their size and scope (especially in the case of voids and filaments), the reliance on N-body simulations to test theory against observation makes them less desirable and technologically more difficult (but still essential) tests when compared to analytical solutions of linear theory. This is not to say they can be ignored. On the contrary, in the current state of Δ domination, density contrasts continue to grow, breaking down the use of not only linear- perturbation theory but all perturbative treatments outside of direct simulation; thus, a theory that connects fundamental physics to the large-scale structure is still paramount. However, the breaking down of linear theory is time and space-dependent, more minor scales of space break down before larger scales, so as long as one observes a large enough scale at a high enough redshift, one can still test perturbations against their linear theory predictions.

Which begs the question; How do we test linear theory against observation? As equation 1.64 shows the growth of perturbations over time, one may wish to test individual perturbations by predicting and then observing $\delta(x,t)$ where x and t are any given point in space and time. However, this approach has severe limitations, as we can only receive observable signals within our light cone. One does not have all available x,t to observe, let alone the initial conditions for any given x. Furthermore, determining $\delta(x,t)$ would be a technologically massive task because the quantity would contain so much information, essentially giving the power to predict the position of every galaxy throughout time. Thus a test of this sought is not only impossible but incredibly inefficient (Piattella, 2018).

Alternatively, we can treat the perturbations as random variables (the collection of random variables being the random field) where there initial value of any point is randomly based on probability distribution. Statistical information can then be inferred about this field, such as expectation value, variance, and the equations' statistical quantities can predict. We can define a random field as $G(x)$ where the probability of the field being a certain value g_1 at any one place x_1 is defined as

$$p(G(x_1) = g_1) = p_1(g_1)dg_1. \tag{1.70}$$

As we move under the assumption that the universe is homogeneous on the largest scale, we can assume that the probability function for one point in space is the same as for another. Thus we can write the expectation field as independent of x,

$$\langle G \rangle = \int_{\Omega} g p(g) dg, \tag{1.71}$$

where Ω is the ensemble, thus integrating all possible states. Of course, in the case of our universe, we do not have access to every possible state as we only have one single realisation of our universe, so instead, an estimator would be used that takes a spatial average of a given volume V

$$\bar{G} = \frac{1}{V} \int_{V} d^3 x G(x). \tag{1.72}$$

It can be proven that the error on this estimator, when compared to the actual expectation value, will be, on average, 0 and that the variance on this error goes to 0 as $V \to \infty$ (Moore, 2015). However, as stated previously, the reason we cannot access $\delta(x,t)$ we do not have access to the entire universe, only a finite observable cone; thus, there will always be some possible error in our calculation of $\langle G \rangle$ known as cosmic variance (Moster et al., 2011). This is the ideal situation as most modern studies are dominated by data variance of having observation areas much smaller than the light cone itself. An interesting statistic to consider is the probability of $G(x_1)$ and $G(x_2)$ having the values g_1 and g_2 which is given as $p_{12}(g_1, g_2) dg_1 dg_2$ one can then use this probability to define the 2-point correlation function $\xi(x_1, x_2)$

$$\xi(x_1, x_2) = \langle G(x_1) G(x_2) \rangle = \int_{\Omega} g_1 g_2 p_{12}(g_1, g_2) dg_1 dg_2. \tag{1.73}$$

If we consider the field to be homogeneous and isotropic, the 2-point correlation function can simplify to $\xi(x_1, x_2) = \xi(r_{12})$ where $r_{12} = |x_1 - x_2|$ as the function is only dependent on the distance between 2 points, not the points themselves or the direction of the difference. As a physical interpretation, consider a random field that's purely Poissonian, so the probability at each point is independent of all other points ie.$p_{12}(g_1, g_2) = p_1(g_1) p_2(g_2)$ under these conditions $\xi(r_{12}) = \langle G \rangle$ thus any deviation from the mean represents a non-Poisson random field, one of which points correlate with other points. Equally, we can say the probability of finding a pair of galaxies at a separation of r_{12} in volume sizes dV_1, dV_2 (see Figure 1.9) is given as $dP = \bar{n}^2 (1 + \xi(r_{12})) dV_1 dV_2$ where $\bar{n}^2$ is the mean galaxy density. For example, if the random field is galaxies, one may expect a nearby positive correlation as gravity will bring galaxies closer together. With this in mind, the 2-point correlation function is a great statistic for measuring the effects of gravity on our perturbations.

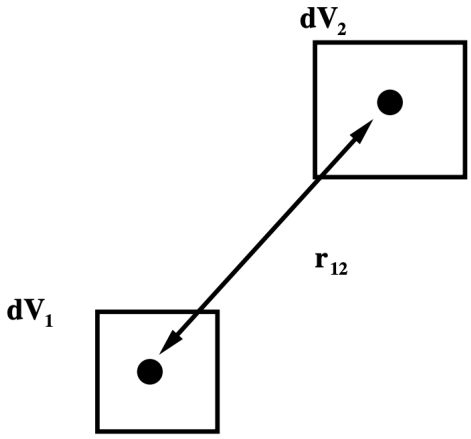


Figure 1.9: An illustration of the 2 point correlation function which is the excess probability finding 2 galaxies r_{12} apart in volume elements dV_1 and dV_2 compared to a random distribution (Reid et al., 2019).

We can also Fourier transform our random field to the following:

$$G(x) = \int \frac{d^3k}{(2\pi)^3} \tilde{G}(k) e^{ikx}, \tag{1.74}$$

$$\tilde{G}(k) = \int d^3x G(x) e^{-ikx}, \tag{1.75}$$

where k is the wavenumber. This allows us to consider the 2-point correlation function of the FT field

$$\langle \tilde{G}(k)\tilde{G}^*(k')\rangle = \int d^3x e^{-i(k-k')x} \int d^3z \xi(z) e^{ik'z}, \tag{1.76}$$

where $z = x' - x$. By introducing the Dirac Delta function δ_D, we can define the power spectrum P(k),

$$\langle \tilde{G}(k)\tilde{G}^*(k')\rangle = (2\pi)^3 \delta_D^{(3)}(k-k')P(k). \tag{1.77}$$

$$P(k) = \int d^3x \xi(x) e^{ikx} \tag{1.78}$$

Thus the power spectrum is the Fourier transform of the correlation function. Making this change is beneficial. If our random field is the matter perturbations and we consider those perturbations to be Gaussian in distribution, then the field modes δ_k grow independently in equation 1.64. We make predictions of the matter power spectrum at different redshifts $P_\delta(k,z)$ as

$$P_\delta(k,z) = T(k,z)P_\delta(k), \tag{1.79}$$

where $T(k,z)$ is the transfer function that describes equation 1.64 and $P_\delta(k)$ the initial conditions. The remaining issue is the one mentioned above. One can not observe the power spectrum directly as it's an ensemble average; instead, one relies on the correlation function (spatial average) in a finite volume $V = L^3$ of the survey leading to a degree of cosmic variance. However, the amount is negligible as long as the scale k being observed is much smaller than the survey size $kL \gg 1$.

As the use of power spectra as an apt statistic to describe the clustering behaviour of perturbations relies on the assumption that the field is Gaussian, it is worth discussing how robust such an assumption is. It is already clear that this assumption does not hold for minor scales due to non-linear evolution, as mentioned above. So any Gaussianity is assumed to be limited and tested at large scales where linear-perturbation theory still holds. Such an example of a test was carried out using data from the IRAS QDOT survey (Efstathiou et al., 1992) which contained 1824 galaxies at a radius of $20\,h^{-1}\text{Mpc} < R < 500\,h^{-1}\text{Mpc}$ from the earth, by covering 73.9% of the sky this would seem a large enough scale. By applying the FKP estimator (an early estimation for the matter power spectrum) to the survey data, the analysis showed that matter density modes had power values that are independently exponentially distributed (a prediction of gaussianity) up to powers 10 times the mean value. They found no traces of non-Gaussian behaviour at all (Feldman et al., 1993).

That said, the result has had pushback (or, more precisely, the conclusions interpreted from the result). Firstly on the theoretical side, it has been argued that measuring the power of individual modes does constitute a complete test of gaussianity and that higher-order k-space correlation would be required to claim model independence (Peacock and Dodds, 1994). However, there have been observations that contradict the result entirely. For example, a study of 4 pencil beam surveys (small sky cover but high redshift) containing a well-sampled distribution of galaxies up to $2000\,h^{-1}\text{Mpc}$ also looked at individual modes and found a periodicity in the galaxy distribution every $128\,h^{-1}\text{Mpc}$, a sure sign of non-gaussianity (Broadhurst et al., 1990). However, re-analysis of the same data set has suggested that if one includes the effects of realistic-survey geometry, the noise level increases to the point where the result is insignificant (Kaiser and Peacock, 1991). Other tests also exist, such as the one-point distribution of the velocity field (Nusser and Dekel, 1993) and the topology of isodensity surfaces (Hamilton et al., 1986), but so far, none have produced evidence against Gaussian (Peacock and Dodds, 1994). This may bring up the philosophical point that when testing for non-gaussianity, the absence of evidence is not evidence of absence. The argument has also been made that due to the limited size and scope of past surveys, non-Gaussian fields may appear Gaussian due to the central limit theorem alone, questioning the validity of their use (Scherrer, 1992).

Outside of the Gaussian field assumption and the non-linear evolution of minor scales, 2 other significant effects must be considered when trying to observe the matter power spectrum. Firstly by 'matter', we include baryonic and Dark combined (although hugely dominated by the dark matter component), but as shown earlier, we cannot observe dark matter directly, so instead, we must rely on observation. As they are believed to form in the centre of dark matter halos, they should be excellent tracers for the underlying matter. However, this relationship is not one-to-one, and the need for galaxy bias b_g is needed to link the galaxy and matter power spectra,

$$P_g(k) = b_g^2 P_\delta(l), \tag{1.80}$$

where P_g and P_δ are the galaxy and matter power spectrum, respectively, the galaxy bias b_g is the first term in a series expansion of the galaxy overdensity δ_g, but one can use just the linear term to describe large-scale mass power spectra. For a complete description, the galaxy bias would have to be a function of redshift and scale. Although models exist (purely analytical and based on simulation), they still break down on minor scales. This is to our detriment as many cosmological parameters in the context of galaxy clustering are degenerate with galaxy bias (Samuroff, 2017). The issue may even compound with the possible inclusion of non-Gaussian initial conditions where the other scale dependencies widen the gap between a galaxy cluster and matter that cannot be explained through non-linear evolution alone (Desjacques et al., 2016).

Another effect that has to be considered is redshift-space distortions. When one is not taking a 2-D projection of the sky but instead dealing with a 3-D data set with some 'depth', redshift is usually used as the radial coordinate. However, as we saw in 1.3.1, the redshift one observes a combination of expansion and peculiar velocity. Thus any peculiar velocity in observed galaxies will distort the observed matter power spectrum even in the generous case of no noise data (Peacock and Dodds, 1994). The effect on data can be split into 2 distinct categories. The first is large scales, known as the anisotropic effect (Kaiser, 1987) where mass from low-density regions moves towards high-density sheets. If those sheets lie near the plane in the sky, the density contrast between the 2 regions will be enhanced in redshift space. This effect on the model density is summarised as

$$\delta_k \to b\delta_k(1 + f\mu^2/b), \tag{1.81}$$

where b is a constant, f is a velocity suppression factor, and μ is the cosine of the angle between the wavevector and the line of sight. The second effect refers to small scales where visualised motion smears information in the radial component reducing power in the power

spectra. This effect is hard to treat precisely, but simplified models exist (Peacock, 2008). Although efforts have been made to combine the 2 effects by multiplying the 2 correction factors (Fisher et al., 1993) this brings mathematical difficulties as both effects interfere with each other, leading to a new correction altogether.

With an understanding of the 2-point correlation function and thus the matter power-spectrum, one can finally mention this thesis's other critical cosmological parameter outside of Ω_m. To track the growth of structure, it may be necessary to not only know the density of matter in the universe but also how 'clumpy' that matter becomes. In other words, one may wish to find the variance in the matter perturbations $\langle \delta^2(r,t) \rangle$. This same variance can be extracted directly from the power spectrum,

$$\langle \delta^2(r,t) \rangle = \int_0^\infty \frac{dk}{k} \Delta^2(k), \tag{1.82}$$

where $\Delta(k) = k^3 P(k)/2\pi^2$ is a dimensionless version of the power spectrum. $\Delta(k)$ can be understood as the measure of density fluctuations on the scale of k^{-1}. In this current state, the perturbation variance is not particularly used for 2 reasons. Firstly one would never be able to observe density in an individual point (we always observe over a volume no matter how small that volume is), so we could not have an observation to use for comparison. Secondly (although the points are intrinsically connected), the integral of the equation 1.82 does not converge with increasing k (or decreasing scales), so the variance could never be found. Instead, one needs to force convergence in the integral by averaging regions over a volume, usually a sphere of radius $r\,h^{-1}$ Mpc. This effective variance is defined as

$$\sigma_R = \int_0^\infty \frac{dk}{k} \Delta^2(k) \tilde{W}(kR)^2, \tag{1.83}$$

where $\tilde{W}(kR) = 3[\sin(kR) - (kR)\cos(kR)]/(kR)^3$ is the Fourier transform of the top hat function

$$W_R(y) = \begin{cases} 1/V_R & y < R \\ 0 & y > R \end{cases}. \tag{1.84}$$

Convergence is obtained as $\tilde{W}(kR)^2$ decays as $(kR)^6$. Although any sized window function can be used, the choice of the value of R is important when it comes to usefulness. Too large and small-scale effects are potentially smoothed out. Too small, and the reliance on unreachable observables is impractical. Typically the value used is $R = 8\,h^{-1}$ Mpc, roughly the size of a sphere of material required to form a galaxy cluster which seems then a good scale to choose as galaxy clusters are one of our most reliant observable (Lahav and Liddle, 2014). The other reason is that non-linear cosmology only dominates the variance when

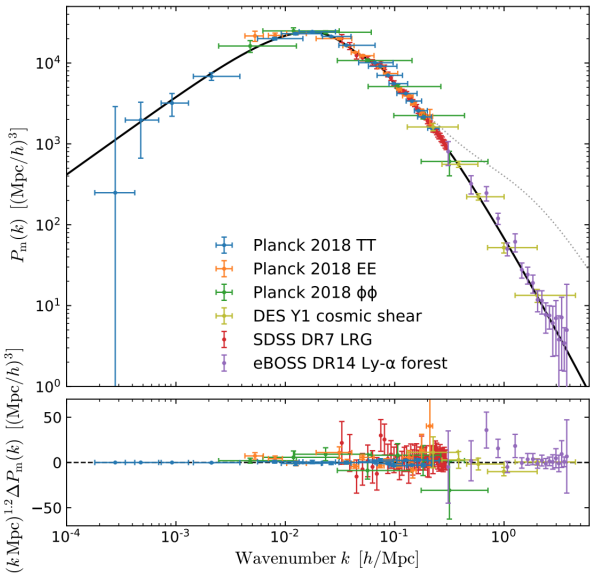


Figure 1.10: *Top*: Data points inferring the matter power spectrum at z = 0 using a combination of different probes, including PLANCK CMB Data, SDSS galaxy clustering, SDSS $Ly-\alpha$ and DES Cosmic shear. The solid black line represents the theoretical prediction using best-fit Planck 2018 parameter values in the ΛCDM model. The dotted line represents the theoretical prediction with the inclusion of non-linear effects. *Bottom*: Deviations between data and theoretical prediction (Chabanier et al., 2019).

$\sigma_R^2 = 1$ is around $R \sim 8$. We define this particular variable σ_8, the true value of σ_8 is one of the main tensions still existing within ΛCDM, but this will be covered in Chapter 2.

1.4.5 Probes

Seeing that as long as we stay within linear scales, power-spectra provides all the information for a Gaussian system. With this in mind, the standard ΛCDM test compares observations of the matter power spectrum to theoretical predictions. However, power spectra are a function of scale and different observables probe at different scales. As one can see from Figure 1.10 (Chabanier et al., 2019) one covers a large amount of scales ($2 \times 10^{-4}\, h\,\mathrm{Mpc}^{-1} < k < 2\, h\,\mathrm{Mpc}^{-1}$). At the largest scale, anisotropies of the microwave background (CMB). On intermediate scales ($\sim 10^1$ Mpc), Cosmic Shear and galaxy clustering data are used, while on the smallest scales ($\sim 10^0$ Mpc), the Lyman alpha (Lyα) forest is used. Different observations allow one to also probe at different redshifts, in this case, $z \sim 10^3$ for CMB, $2.2 < z < 4.6$ for Lyα, $0.2 < z < 1.3$ for Cosmic Shear, and $z \sim 0.35$. Such differences in redshift allow for the testing of ΛCDM as one should have consistency in result no matter the redshift.

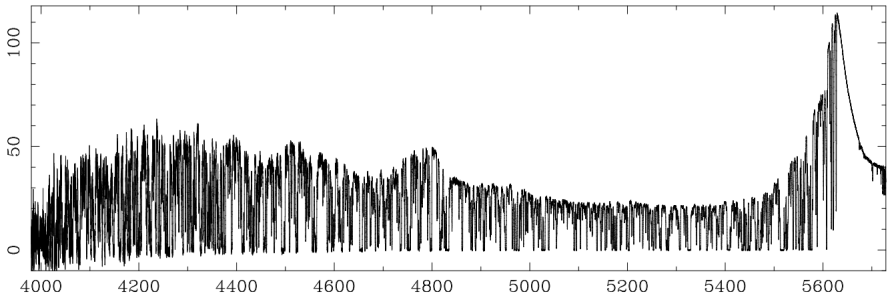


Figure 1.11: The high-resolution spectrum of QSO $1422+23$ at an emission redshift of $z_{em} = 3.62$ shows the many absorption lines of creating the Lyα forest,(Rauch, 1998).

The Lyα Forrest refers to observations of the intergalactic medium (IGM), the gaseous environment surrounding galaxies (Spitzer and L., 1956) (this is the same environment that holds the majority of Byronic matter of the universe as seen with the bullet cluster above). The IGM would mainly contain the most abundant elements in the universe, Hydrogen and Helium, which would be ionised using the radiation of nearby galactic stars. There would, however, be a fraction of the neutral gas (through the capturing of electrons in the ionised plasma) (Gunn and Peterson, 1965). This fraction of neutral gas can be observed as neutral Hydrogen absorbs ultraviolet light (at a photon wavelength of $\lambda = 1215\,\text{Å}$). Which at $2.2 < z < 4.6$ redshifts into observable visible light (3900 - 7000). As not all absorption would take place simultaneously (as the IGM exists across redshifts), the wavelength of the absorption would also vary, leading to 1000s of small absorption lines only a few angstroms wide, leading to the name Lyman α Forest as the absorption lines mimic a dense Forrest (Weymann et al., 1981) (as can be seen from the spectral lines of the quasistellar object QSO $1422+23$ in Figure 1.11 (Rauch, 1998).

The forest is of such cosmological use as it can give us an idea of the density of baryonic matter within the IGM. Moreover, this same IGM has been shown to have linear perturbations in density, meaning that observable properties can be predicted from minimal initial conditions without the need for non-linear physics. In this sense, the forest acts as a tracer is a tracer for the matter (and not just baryonic matter) density of the universe. Through this usefulness, Lyα has provided constraints on dark matter and dark energy parameters through measurements of Baryonic Acoustic Oscillations (Bautista et al., 2017) which are a minor peak in the matter 2-point correlation function caused by a delayed decoupling of baryonic matter from photons compared to dark matter, this minor peak appears as oscillations in Fourier-space (power spectra). It has also probed the structure formation of neutrino (Yèche et al., 2017) masses but most importantly for this context. Finally, it provided observations of the matter power spectrum. The observations present in Figure 1.10 were produced through

the BOSS and eBOSS surveys (Abolfathi et al., 2018) which used 43,751 quasars (highly luminous centres of the galaxy) spectra. With thorough knowledge of the redshifts ,of the quasars, a line of sight flux power spectra $P_{F1D}(k)$ (Chabanier et al., 2018) can be produced, which can be directly related to the matter power spectra,

$$P_m(k) = \frac{1}{b^2}(-\frac{2\pi}{k}\frac{dP_{F1D}}{dk}), \tag{1.85}$$

where b is a scale and redshift dependent bias that depends on the model parameters (most notably Ω_m and σ_8).

At a redshift of $z = 1100$, the cosmic microwave background observations represent the earliest images we have of our universe. First observed by 2 radio astronomers (Arno Penzias and Robert Wilson) as background noise from their radio telescope, equating to an antenna temperature of 3.5K (Penzias and Wilson, 1965). The noise was, in fact, photons from the point of recombination. Before $z = 1100$, the universe was a hot plasma as the temperature was too high to create Hydrogen. The universe was 'opaque' in that photons could not travel far without interacting with electrons and changing direction (Thomson Scattering (Bashkansky and Reintjes, 2004)). In this state, photons and baryonic matter were 'coupled' (shared similar temperatures). However as the universe expanded the temperature of the universe cooled (as $E(z) \propto T(z) \propto 1/a(z)$). At about 3000 K, protons and electrons combined to produce Hydrogen and the number of free electrons decreased, and photons were not energetically high enough to interact, ionise and reverse the Hydrogen making process. Photons and Byronic matter had decoupled, creating a transparent universe in which today we can observe photons as there were at that point.

As was the case previously with the Lyman-alpha forest, the cosmic microwave background opened the door to observations untouched by non-linear physics. At its most basic cosmic radiation was strong evidence for the big bang model, that we started in the hot dense initial state and then expanded and cooled cosmic radiation was an initial prediction of the model before its eventual discovery (Gamow, 1948). The data of interest concerning the CMB are perturbations in the temperature field. Although fairly uniform ($\delta T/T \sim 10^{-5}$ (Bucher, 2015)), the small differences have huge cosmological implications. One such example is the dipole. On the scale of 180 degrees, a perturbation in temperature of $\delta T/T \sim 10^{-3}$ is observed. At first glance, such a large-scale perturbation would not be expected given the lack of large-scale structure, but instead, it can be explained as blue/red shifting due to the peculiar motions of the earth itself (Durrer, 2015). This same observation has lead values for the peculiar velocity of the solar system that are consistent, which is much lower than redshift observations (Aghanim et al., 2014). In the context of the *LambdaCDM* model,

the Temperature power spectrum is precious. Because one observes the whole sky instead of a section, one cannot use normal Fourier analysis in 2-dimensions. Instead, we refer to the temperature fluctuations in the direction of the sky $\hat{\mathbf{n}}$ as $\Theta(\hat{\mathbf{n}}) = \Delta T/T$. These same perturbations on a sphere can then be transformed into multipole moments (the angular equivalent of Fourier waves),

$$\Theta_{\ell m} = \int d\hat{\mathbf{n}} Y^*_{\ell m}(\hat{\mathbf{n}})\Theta(\hat{\mathbf{n}}), \tag{1.86}$$

where $\Theta_{\ell m}$ is the transformation, $Y^*_{\ell m}(\hat{\mathbf{n}})$ the multipole moments. The angular wavelength can be related as $\theta \sim \pi/\ell$. Assuming the perturbation field is Gaussian, it can be completely characterised through the angular power spectrum C_ℓ,

$$\langle \Theta^*_{\ell m}\Theta_{\ell' m'}\rangle = \delta_{\ell\ell'}\delta_{mm'}. \tag{1.87}$$

Although, as one can see from the CMB spectrum in Figure 1.12 the CMB power spectrum is usually displayed in its unitless form,

$$\Delta^2_T = \frac{\ell(\ell+1)}{2\pi} C_\ell T^2. \tag{1.88}$$

Figure 1.12 also illustrates just how sensitive the shape of the CMB power spectrum is to cosmological parameters. A case in point is a) which refers to variations in Ω_{tot} or the universe's curvature. The peak in the angular power-spectrum ℓ_{peak} is directly linked to the angular size of the sound horizon, a known value on the angular scale of 1 degree. Any deviations from this value would be evidence of curvature. Figure 1.12a) shows horizontal shifts in the first peak. This is due to the usual spots in the CMB map, which correspond to the peaks appearing larger or smaller depending on if the light had bent inwards or outwards depending on its closed or open curvature $\ell_{\text{peak}} \propto 200/\Omega^{1/2}$. These same CMB spots are some of the best evidence we have for the flatness of our universe (de Bernardis et al., 2000). Similarly, one can see the different impact matter and baryon densities (Ω_m, Ω_b) have on the cosmic microwave background power spectra. 1.12 d) shows that as the matter density increases, the amplitude of the peaks relative to the large-scale plateau decreases. This is because large-scale modes only began to grow in matter domination. Decreasing the matter density enhances the small-scale modes that entered the horizon during radiation domination (Benjamin, 2018). As for the baryon density in 1.12 c), the relative peaks are affected by their size, with odd peaks growing with Ω_b relative to even peaks. This is due to an effect called 'baryon loading'. The coupled photons and baryons oscillate in and out of the dark matter potential wells due to gravity and radiation pressure. However, as baryons contain

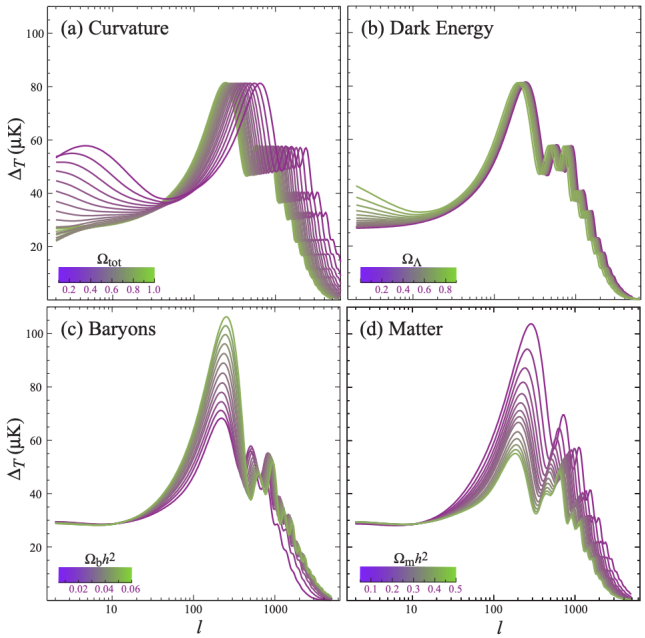


Figure 1.12: Sensitivity of the CMB power spectrum under 4 cosmological parameters a) the curvature as indicated by Ω_{tot} b)energy density of dark energy Ω_Λ c) baryon density $\Omega_{\text{b}}h^2$ and d) matter density $\Omega_{\text{m}}h^2$. The variations take place over a fiducial model of $\Omega_{\text{tot}} = 1$, $\Omega_\Lambda = 0.65$, $\Omega_{\text{b}}h^2 = 0.02$, $\Omega_{\text{m}}h^2 = 0.147$ (Hu and Dodelson, 2002).

mass, they also contribute to the gravitational effect so that the dark matter and baryon caused compression (the odd peaks) are more significant than the photon caused expansions (even peaks). This effect would only become more prominent through the increase in Ω_b. The lack of variation in cosmic microwave background power spectra due to changes in dark energy density is also evidence for our picture of the universe. At such high redshifts, the dark energy density is low compared to radiation and matter so it would have minimal effect. However, as expansion continues and matter and radiation energy decrease, the constant dark energy becomes more prominent, thus why it has more of an effect on the large-scale later redshift modes. The cosmic microwave background power spectra can directly impact the matter power spectrum because the CMB maps out the dark matter potential wells from which matter perturbations grow. As one can see from Figure 1.10 our observations of the matter power spectra are still entirely dependent on observations of the CMB from Planck temperature, polarisation and lensing reconstruction power spectra. Most notably, in 2 regions, the largest scales, which should still have the same linear growth as the original potentials wells and the baryon acoustic oscillations, which were multiple peaks in the CMB, now appear dampened wiggles in the matter power spectra.

Another and probably more apparent probe is the study of galaxy clusters. Galaxy clusters are some of the most significant objects in the universe, containing hundreds and possibly thousands of galaxies within them. Such an example is the Virgo Cluster lying 50 million lightyears away and containing over a thousand galaxies. It is the closest galaxy cluster to the eX-ray By largest, we mean most massive gravitational bound systems, with some galaxy clusters as high as $10^{15} M_{\odot}$.

The advantage of high masses is that they contain many objects (such as the billions of stars within each of their hundreds of galaxies) that emit light for observation across many wavelengths. Such an example is the light in the optical and near-infrared (stars) which has been used to study the galaxy luminosity function (the number distribution of galaxies at certain luminosities), which gives insight into galaxy formation and evolution (Popesso et al., 2005). The other much larger photon emitting source of galaxy clusters is in the form of extremely hot defused intergalactic gas. The compression due to the gravitational potential of the cluster courses the gas to rise to density and temperature, emitting X-ray radiation (Forman et al., 1978). This is the same x-ray emission used as evidence for dark matter, especially in cases such as the bullet cluster. However, this emission is not the only use of the IGM and not the only way we can detect it within galaxy clusters. Photons from the last scattering surface (CMB) interact with the case before reaching the earth. These interactions lead to an increase in photon frequency. One can directly observe this phenomenon named the Sunyaev-Zel'dovich (SZ) effect by measuring the intensity in CMB photon frequency and noticing a deficit and surplus of intensity below and above the interaction frequency $\mu = 217\,\mathrm{GHz}$ (Birkinshaw et al., 1984). Measuring the size of the SZ signal is an essential observable being highly linked to the cluster's mass and correlates with the gas mass and temperature. A much smaller secondary SZ effect is caused by the peculiar motion of the IGM known as the Kinetic-SZ effect, or the Ostriker-Vishniac effect, which is instrumental for understanding cluster motion (Ostriker et al., 1986).

This wealth and diversity in galaxy cluster signals allow for multiple probes of modern cosmology. Simulations have shown that with a fixed, known number density of clusters and different redshifts, one can constrain Ω_{M} and σ_8 (Borgani and Guzzo, 2001). That is because clusters form in dark matter halos, requiring initial perturbations on a comoving scale of $\sim 10\,h^{-1}\,\mathrm{Mpc}$, the larger σ_8 is, the more perturbations of this size can exist. One can also make the inference that because galaxy clusters require such large comoving volumes (or an initial perturbation of a certain scale) to collapse, then baryon fraction $f_{\mathrm{b}} = \Omega_{\mathrm{b}}/\Omega_{\mathrm{M}}$ or this clusters should be constant as the dark matter and baryon matter within would not differ much from the cosmic average. Thus by thorough knowledge of the cluster gas fraction through x-ray observation and knowledge of Ω_{b} through primordial nucleosynthesis models,

one can directly obtain Ω_{M} (Bridle and King, 2007). One can also use galaxy clusters as 'standard candles'. These objects have a known distance in the universe and can be useful for determining H_0. This is possible due to the SZ signal, and the X-ray observations give us insights into the thermal properties of the IGM. Provided the assumption that the cluster is spherically symmetric, one can estimate the angular diameter distance of the cluster

$$D_A \propto \frac{(\Delta T_{\mathrm{CMB}})^2 \Lambda_{ee}}{S_X T_e^2}, \tag{1.89}$$

where ΔT_{CMB} is the temperature change due to the SZ effect, S_X the X-ray flux, Λ_{ee} the X-ray cooling function and T_e the electron temperature (Bonamente et al., 2006).

In the case of the matter power spectrum in Figure 1.10 instead of looking at individual galaxy clusters, one can observe the overall galaxy clustering. In this case, a subset containing 110576 Luminous Red Galaxies from the SDSS galaxy survey (Reid et al., 2010) was used to recover the halo density field (as galaxies are dark matter halo tracers) and thus the halo density power spectra, which can be related to the underlying mass power spectrum through the appropriate bias model.

1.5 Summary

From the chapter above, one can see that from various significant observations, we now have a working model for the universe we are a part of. A working model that is only reliant on a small set of logical assumptions (Perivolaropoulos and Skara, 2021) :

- Gravity and the dynamics of space-time are correctly described by the Theory of General Relativity.
- That on cosmological scales ($>$ 100Mpc), the universe is statistically homogeneous and isotropic, which under GR leads to the special part of our universe being described through the Friedmann-Lema^ıtre Roberson-Walker (FLRW) metric.
- The density components of the universe consist of matter in the form of cold dark matter and baryons, radiation in the form of neutrinos and photons and dark energy, which is responsible for the universe's expansion.
- The large-scale structure of today's universe can be described through gravitational dynamics taking place on primordial Gaussian perturbations caused by initial quantum fluctuations.

Although not mentioned above in any detail, the last important assumption is that before the CMB, the universe went through a phase of rapid accelerated expansion (inflation), and it is this phase that bridges the gap between quantum fluctuation and Gaussian perturbation. Combining these assumptions leads to a model that requires only a small number of independent parameters, which include baryon and matter densities (Ω_b,Ω_M), the Hubble constant h as well as some less explored (within this thesis) parameters such as the optical depth τ and amplitude A_s. These parameters have led to a model (ΛCDM) that has proved extremely successful in describing our universe. Not only has it survived more than a decade's worth of precision tests (Bull et al., 2015) but today, our most modern and precise measurements of the CMB (Collaboration et al., 2018) have found that the ΛCDM model can correctly describe 'the cosmological information in over a billion map pixels with just six parameters' (Aghanim et al., 2020) with 5 of the 6 parameters simultaneously measured to sub-per cent precision. No compelling evidence has yet been shown to add extensions to the base-ΛCDM model and even less for a new model entirely. These successes are not limited to early redshift CMB data, as we saw earlier with Fig. 1.10 observations on multiple scales that continue to late redshift matter power spectra in agreement with the spectrum predicted by the parameters of Plank.

With that said, cosmological physics is far from complete. Weaknesses still exist in the assumptions required for the ΛCDM model. Some, such as the nature of cold dark matter, are hoped to be rectified through further experimentation and observation. Others, such as an understanding of dark energy or the marriage of GR to the other most successful theory in physics (Quantum Field Theory) may require new theoretical innovations. Even if we ignore those assumption issues, minor problems still exist within ΛCDM, most prominently the tension between late and early observations of the Hubble constant H_0. This is not the only tension. Other issues also exist where measurements are not entirely aligned as expected. Some of these discrepancies are briefly listed below.

- The CMB cold spot is huge and not consistent with the predicted gaussianity of ΛCDM (Cruz et al., 2004)
- The peak position of the Baryon Acoustic Oscillation as observed by Lyα and the CMB differ by 2.5-3σ (Aubourg et al., 2014)
- Recent CMB observations show a 3.4σ preference for a closed universe, however when combined with BAO observations a flatter universe $\Omega_K = 0.0008 \pm 0.0019$ is preferred (Collaboration et al., 2018)

- On small scales ($< 100\,\text{kpc}$), inconsistencies arise from our current dark matter model. These have already been covered above but include the core-cusp problem and the missing satellite problem

- The lithium abundance calculated from observations of absorption lines from metal-poor halo-stars in a milky way leads to a value of $Li/H = (1.6 \pm 0.3) \times 10^{-10}$ (Group et al., 2020) which is 3.5 times smaller than the theoretical prediction of $Li/H = (5.62 \pm 0.25) \times 10^{-10}$ a tension of over 5σ (Pitrou et al., 2018)

- Using quasars to extend Hubble diagrams to higher redshifts ($z < 5.5$) reveals tension in the observed distance modulus-redshift relation against flat ΛCDM to be at 4σ (Risaliti and Lusso, 2018) a tension that is made worse through the inclusion of high redshifts Gamma-ray bursts and Pantheons (Lusso et al., 2019)

- Through torsion balance experiments, a small (sub-millimetre) wavelength oscillating force has been observed at a 2σ level. This is in direct conflict with the small-scale forces predicted by Newtonian gravity (Perivolaropoulos, 2016)

- An interacting pair of high mass high redshift galaxy clusters known as El Gordo (ACT-CL J0102-4915) (Menanteau et al., 2011) have a high relative velocity which indicates that they were formed earlier than is possible within ΛCDM to a 6.16σ confidence (Asencio et al., 2020)

Other than the ones mentioned above, other tension and anomalies exist, such as anomalously strong ISW effects, a lack of large-angle CMB temperature correlations, Age of the Universe issues, and many more. For a deeper insight into all tensions covered, please see the review by Perivolaropoulos and Skara (2021). The investigation of these tensions can only illuminate our understanding of the universe through the tensions evaporating under a better understanding of systematic errors and more data or through the need for new theoretical modelling. There has, however, been one significant omission from the topic so far. A tension like H_0 separates the value of a cosmological parameter between low and high redshifts. That tension is σ_8, whose primary cosmological probe has been omitted so far and is the focus of the next chapter; Weak Lensing.

Chapter 2

Weak Lensing

2.1 From Newton to Einstein

Weak lensing is connected to a phenomenon known as gravitational lensing, in which photons' trajectories are changed and bent in the presence of mass. These trajectory changes lead to distortions in the images we observe, much like a lens would distort an image. It is these distortions that give insight into the masses that produce them.

Although widely linked to General Relativity, the bending of space-time, and Einstein's interpretation of gravity, one can still understand how light could bend through a Newtonian approximation. Newton himself questioned whether his equations of gravity would act on the light in the same way they act on bodies of mass (Newton, 1704). At first, this seems problematic, if one looks at the equation for the gravitational force given as

$$F_g = \frac{Gm_1m_2}{r^2}, \tag{2.1}$$

where G is the gravitational constant, and r is the distance between 2 bodies of masses m_1 and m_2. Light is presumed to have a mass of 0, which under the equation 2.1would lead to no force interaction. However, if one considers the acceleration of photon a_1, by using Newton's 2nd law of motion that $F = ma$, one can determine the acceleration of the photon due to gravity as

$$a_1 = \frac{Gm_2}{r^2}. \tag{2.2}$$

An acceleration entirely independent of the photon's original non-existing mass m_1 (a conclusion already reached by Galileo, who observed that as long as gravity is the only interacting force, all bodies, regardless of their mass with the same initial velocities, follow

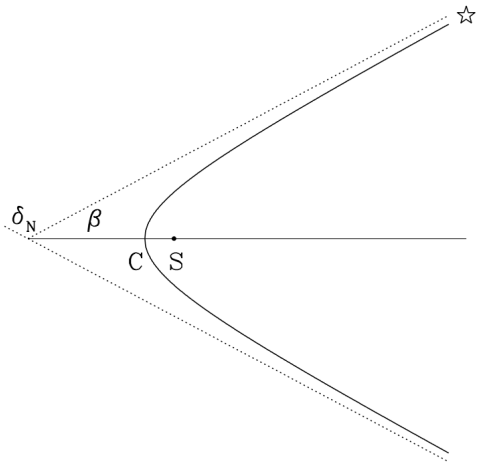


Figure 2.1: A distant star's light ray C is deflected through gravitational interactions with the sun S. δ_N is the Newtonian deflection angle between the light's original and now changed trajectory (Soares, 2005).

the same falling trajectory). So as long as the light is treated as a body (in this case, a photon particle as opposed to a wave), one could predict its trajectory change in Newtonian physics.

The case marks an observable change in trajectory due to how much a light path is bent by the presence of a massive object such as the sun. For example, consider Figure 2.1. Light from a distant star will not follow a straight line trajectory but instead be bent by the presence of the sun S to follow path C. How much this path has bent is given by the deflection angle

$$\delta_N = \pi - 2\beta, \tag{2.3}$$

where β is intrinsically linked to the eccentricity ε of the trajectories hyperbole as $\beta = \cos^{-1}(1/\eta)$. For a normal particle of mass m, energy E and angular momentum L (concerning the sun), the eccentricity is calculable,

$$\begin{aligned} \varepsilon &= (1 + \frac{2EL^2}{G^2 m^3 M_S^2})^{1/2} \\ &= [1 + \frac{(v^2 - 2GM_S/R_S)c^2 R_S}{G^2 M_S^2}]^{1/2}, \end{aligned} \tag{2.4}$$

where M_S and R_S are the mass of the sun and the distance between the sun, the point of measuring the deflection angle and v is the particle's velocity. This equation also assumes the particle's kinetic energy remains constant which is a fair assumption if the velocity magnitude is large). In the case of the photon $v = c$ where $c \gg 1$ so the assumption holds. By assuming

$c^2, \gg 2GM_S/R_S$, one can simplify the photon's deflection angle,

$$\begin{aligned}\delta_N &= \pi - 2\cos^{-1}(\frac{GM_S}{c^2R_S}) \\ &= \frac{2GM_S}{c^2R_S},\end{aligned} \tag{2.5}$$

where the final simplification comes through the Taylor series $\cos^{-1}(x) = \pi/2 - \sin^{-1}(x) \approx \pi/2 - (x + ...)$ where $x = GM_S/c^2R_S = 4.7 \times 10^{-5} \ll 1$. In the case of our sun, the Newtonian deflection angle was $\delta_N = 0.85$ arc seconds (Soldner, 1804).

Considering the equation 2.5 relied exclusively on Newtonian physics (and having to ignore the issue of force interaction on zero mass objects), the result was not far from the truth. Einstein's first attempts at calculating the deflection angle (before General Relativity was fully formed) had a similar conclusion with a deflection angle of $\delta = 0.83$ arc seconds (Einstein, 1911). For a more definitive answer, one requires the use of full General relativity. As with any GR problem, one must start with the relevant metric. In this case, that metric would not be the FRW (this is a small-scale problem dealing with a point mass; thus, expansion and the cosmological principle do not apply). Instead, we use the Schwarzschild Metric, the solution to spherically-symmetric mass distribution (Schwarzschild et al., 1916),

$$ds^2 = (1 - \frac{2GM}{rc^2})c^2dt^2 - \frac{dr^2}{1 - \frac{2GM}{rc^2}} - r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.6}$$

where (r, θ, ϕ) are spherical coordinates starting as the centre of the point mass of mass M.

To understand the path of a photon under such a metric we must take a brief detour. Consider a particle in its inertial frame of reference, from this frame one is observing proper time $d\tau$ and is constantly at rest ($dx = dy = dz = 0$). Under these conditions equation 1.2 is simplified so that $d\tau^2 = ds^2$ (assuming $c = 1$). A particle would follow a path between spacetime points A and B. The proper time that had elapsed between the 2 can be explained through an arbitrary parameter p that increases along the path.

$$\tau_{AB} = \int_A^B d\tau = \int_A^B \frac{d\tau}{dp}dp = \int_A^B L(x^\mu, \dot{x}^\mu)dp \tag{2.7}$$

where $\dot{x}^\mu = dx^\mu/dp$ and

$$L^2 = \left(\frac{d\tau}{dp}\right)^2 = g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu \tag{2.8}$$

Which is the Lagrange function (squared) of the system. The time lapsed by a particle at rest is stationary (at a maximum). As long as this is the case and the points A and B are fixed the

L^2 must satisfy the following equation.

$$\frac{\partial L^2}{\partial x^\mu} - \frac{d}{dp}\left(\frac{\partial L^2}{\partial \dot{x}^\mu}\right) = 0 \tag{2.9}$$

This is known as the Euler-Lagrange and is used to derive equations of motion throughout mechanics. Typically the equation is shown in ∂L form (as opposed to ∂L^2) but computing $\sqrt{g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu}$ is difficult and avoided where possible.

We can now return to the Schwarzschild Metric. By applying equation 2.6 to equation 2.8 we derive the Lagrange function of the system.

$$L^2 = (1 - \frac{2GM}{rc^2})\dot{t}^2 - \frac{\dot{r}^2}{1 - \frac{2GM}{rc^2}} - r^2(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2) \tag{2.10}$$

Applying this Lagrange to equation 2.9 where $dp = d\tau$ and treating each coordinate in $x^\mu = (r,t,\theta,\phi)$ separately leads to an expression for the conservation of energy ($(1 - 2GM/rc^2)dt/d\tau = \text{constant} = k$) and angular momentum ($r^2 d\phi/d\tau = \text{constant} = h$) but most importantly it leads to the orbit equation (a differential equation showing how distance from the point mass changes along a straight line),

$$\dot{r} + \frac{h^2}{r^2} - \frac{2GM}{r} - \frac{2GMh^2}{r^3c^2} = c^2(k-1) = \text{constant}, \tag{2.11}$$

, which solved for the case of a light path which has $L = ds = 0$ and gives the following solution

$$\frac{1}{r} = \frac{\sin\phi}{R} + \frac{3GM}{2c^2R^2}(1 + \frac{1}{3}\cos 2\phi), \tag{2.12}$$

where R is the distance of the closest approach, consider once again Figure 2.1 we care about the difference in ϕ between where the photon began and where it ends, in other words, far away from the deflecting star, in a system where the sun represents $r = 0$ the initial and final points would be $r \to -\infty$ and $r \to +\infty$ respectively. Under these conditions the ϕ is small so $\sin\phi \simeq \phi$ and $\cos 2\phi \simeq 1$. Applying the r limits to a new simplified equation 2.12 leads to a change in ϕ of $\pi + 4GM/c^2R$. In a normal straight-line course, a π changes in ϕ, so removing that reveals the true deflection angle,

$$\delta_{\text{GR}} = \frac{4GM}{c^2R}. \tag{2.13}$$

This result is precisely twice what was predicted by Newtonian physics when applied to the case of the sun giving a deflection angle of $\delta_{\text{GR}} = 1.7$ arc seconds (Einstein, 1916). This value was confirmed to a 20% error margin during a solar eclipse in 1919 (Dyson et al., 1920)

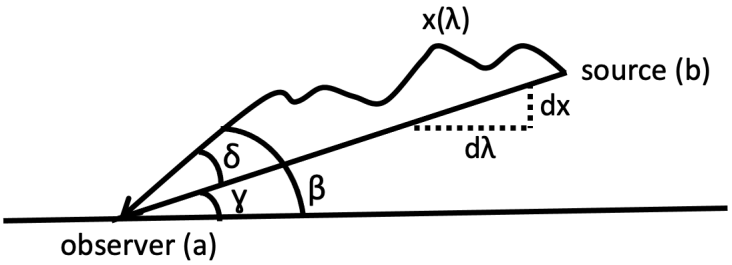


Figure 2.2: A ray path $x(\lambda)$ from source to observer, the straight path represents a light ray without any mass perturbing its path. The angle between observer and source is β and γ for the perturbed and unperturbed path, respectively, the difference between them being the deflection angle δ.

one of the first observational confirmations of GR. More than 7 decades later, the result was reexamined and confirmed with new observations to within 0.02% (Lebach et al., 1995).

2.2 Lens Equations

2.2.1 Fermat's Principle

For Einstein, the discovery of light's path being gravitational bent was nothing more than a visual confirmation of his theory (for more details of this, see the work of Bartelmann and Maturi (2016)). Believing that its effect was too small to be noticed on a cosmological scale. This, however, is not the case. If light can be bent, then masses in the universe have effectively become lenses. It is from the perspective of lenses that we continue going forward. Considering the cosmological case where Newtonian gravitational potentials Φ are small (galaxy clusters have a typical potential of $|\Phi|/c^2 \leq 10^{-5}$) and the mass structure moves slowly relative to the speed of light (said clusters have typical peculiar velocities of $v \leq 600\,\mathrm{km\,s^{-1}}$). Under these conditions, one does not need to calculate the gravitational effect through a Schwarzschild metric. Instead, due to the low potentials being considered (we assume the photons never get close to the point masses), one can use a perturbed Minkowski Metric (tiny changes to otherwise euclidean geometry):

$$ds^2 = -c^2(1+\frac{2\Phi}{c^2})dt^2 + (1-\frac{2\Phi}{c^2})dx^2 \tag{2.14}$$

where $x = (x,y,z)$. Due to photons having $ds^2 = 0$ and through recognising that $1/(1-\frac{2\Phi}{c^2}) \simeq 1+2\Phi/c^2$ when $2\Phi/c^2 \ll 1$ one can rearrange to find what the speed of light is within a weak gravitational field,

$$c' = |\frac{dx}{dt}| = c\sqrt{\frac{1+\frac{2\Phi}{c^2}}{1-\frac{2\Phi}{c^2}}} \approx c(1+\frac{2\Phi}{c^2}). \tag{2.15}$$

In the context of a lens, one can write the refractive index of a gravitational field to be $n = c/c' = 1 - 2\Phi/c^2$. Interpreting the gravitational field as a refractive index allows for using Fermat's principle, which states that a light ray between 2 points will take minimal time if the 2 points are sufficiently close. If not they will take the path of which variations in time t concerning the chosen path are stationary ($dt/dx = 0$) (Born and Wolf, 2019). The principle has been used to explain well known optical phenomena such as Snell's law that under the angles (θ_1, θ_2) of the path light travels from one refractive index to another (n_1, n_2) follows $n_1 \sin\theta_1 = n_2 \sin\theta_2$.

Similarly, we can apply Fermat's principle to determine a generalised statement regarding the photon's deflecting angle. In our context, we consider a path $x(l)$ between 2 points a,b and a refractive index $n(x(l))$ (just as the gravitational potential will vary depending on location) that varies with that path. Under these conditions, we can write Fermat's principle as

$$\delta t = \int_b^a n(x(l))dl = \delta \int_{\lambda_a}^{\lambda_b} n(x(l))|\frac{dx}{d\lambda}|d\lambda = 0, \tag{2.16}$$

where λ is an arbitrary curvature parameter. In this arrangement, our integrand is a Lagrangian function $L(x, \dot{x}, \lambda) = n(x(l))|dx/d\lambda|$ and because we expect the integrand to be minimised we can apply the Euler-Lagrange equation. This leads to the path description given by Fermat's principle

$$\ddot{x} = \nabla_\perp \ln n \approx -\frac{2}{c^2}\nabla_\perp \Phi, \tag{2.17}$$

where $\ln n \approx n - 1$ when n is small. With an understanding of the path, one can relate this to the deflection angle. Consider Figure 2.2 consisting of the path $x(\lambda)$ compared to the unperturbed straight path. We can describe the deflection angle as

$$\delta = \beta - \gamma = \frac{dx}{d\lambda}|_a - \frac{dx}{d\lambda}|_b = \int_a^b \ddot{x}d\lambda. \tag{2.18}$$

Combined with equation 2.17 leads to the direct link between deflection angle and gravitational potential,

$$\delta = -\frac{2}{c^2}\int_{\lambda_a}^{\lambda_b} \nabla_\perp \Phi d\lambda. \tag{2.19}$$

2.2.2 The Born Approximation

A simple method for validating the generalised deflection angle caused by a gravitational potential, as shown in equation 2.19 would be to apply it to our observed scenario of the deflection angle caused by the sun. One typical issue in such cases is that the surrounding gravitational potential constantly changes the light path, making for a highly complex integration. In these circumstances, the Born approximation states that instead of taking the ray's deflected path, we integrate the potential over the straight undeflected path between source and observer. Such a crude simplification is allowed as we deal with deflection angles on the scale of arc seconds. With such minor changes to the angle, one can assume that the deflected and un-deflected path has similar gravitational potentials. As well as simplifying mathematically, the approximation also speeds up calculations computationally as one no longer needs to solve a time-intensive equation such as the Poisson Equation (Petri et al., 2017).

In the case of the deflection angle of the sun, we consider the sun to be a point mass M at the origin of the coordinate system $(x,y,z) = (0,0,0)$, and the light ray follows a straight path unperturbed by the sun's mass (Born approximation) parallel to the z-axis, but with a constant impact parameter of $b = \sqrt{x^2+y^2}$ thus the photon is always $r = \sqrt{x^2+y^2+z^2} = \sqrt{b^2+z^2}$ away from the point mass. So our curve parameter will be z, and the photon travels to and from $z = [-\infty,\infty]$. This requires the perpendicular gradient of the potential,

$$\nabla_{\perp}\Phi = \frac{GM}{r^3}\begin{bmatrix} x \\ y \end{bmatrix}, \tag{2.20}$$

where $\Phi = -GM/r$ the gravitational potential of a point mass. Completing the integral of equation 2.19 reveals the deflection angle in the x and y axis,

$$\delta = \frac{2GM}{c^2}\int_{-\infty}^{\infty}\frac{dz}{\sqrt{b^2+z^2}^{3/2}}\begin{bmatrix} x \\ y \end{bmatrix} = \frac{4GM}{c^2 b}\begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}, \tag{2.21}$$

as $x = \sin\theta$ and $y = \cos\theta$ where θ is the angle between b and the y-axis. To gather the scalar value of the deflection angle, one must take the magnitude,

$$|\delta| = \sqrt{(\frac{4GM}{c^2 b})^2(\sin^2\theta + \cos^2\theta)} = \frac{4GM}{c^2 b}. \tag{2.22}$$

In the case of the impact parameter being the distance of the closest approach ($b = R$), we have a result that is identical to the original result of equation 2.13 (Zaidouni, 2018).

Even with the small-angle justification and the validation of the above example, this is not ample reason to rely on an approximation, primarily if alternatives exist. One such alternative is to deal with the approximation statistically. For weak lensing surveys more concerned with summary statistics (see sub-chapters below), the Born approximation equates to the first-order form of that summary. By using certain second-order terms as well, we are introducing a *Born Correction* (Bernardeau et al., 1996) (Schneider et al., 1998). Another approach is to discretise the problem so that instead of the ray path being a straight line or a continuously evolving line, the ray path consists of multiple straight lines that change direction based on the gravitational potential at particular points (Hilbert et al., 2009). In both cases, the approaches lead to more accurate results; however, based on both analytic and simulated studies, the extra accuracy one gets is negligible (Cuesta-Lazaro et al., 2018). One such example was a simulation of an LSST-like weak lensing survey that found that the born approximation was at least 4 times faster and used half as much memory as tracing the entire path and led to parameter biases on the scale of 10^{-5} which is small compared to other biases in weak lensing. However, if one wished to study non-gaussian aspects of the weak lensing signal, significant biases were introduced, which will have to be examined further with future surveys (Petri et al., 2017).

2.2.3 Lensing Potential

With an understanding of how light paths are deflected by the presence of gravitational potential and a grasp of the appropriate approximations required to calculate them, we can now perform calculations connected to actual images. For example, one may observe distant galaxies. With images, one has multiple light rays from the same source but whose paths differ depending on their original position. These differences lead to distortions in the images we perceive of the original object.

Consider Figure 2.3 which now consists of 2 deflected ray paths not 1 as seen earlier. The source exists at a distance χ from the observer so that in the small-angle approximation, the observer believes the 2 rays are separated by $x_0 = \chi\theta$ (this is for a flat universe, in the case of an open/closed universe $f_K(\chi)$ replaces χ in the equation). This is not their true separation, instead, their separation has been modified so that from the perspective of the mass coursing the distortions a χ' the change in separation is given as $dx_{\text{deflected}} = (\chi - \chi')d\delta$ where $d\delta = -2\nabla_\perp \Phi d\chi/c^2$ as seen in equation 2.19. Thus the true separation of the rays can

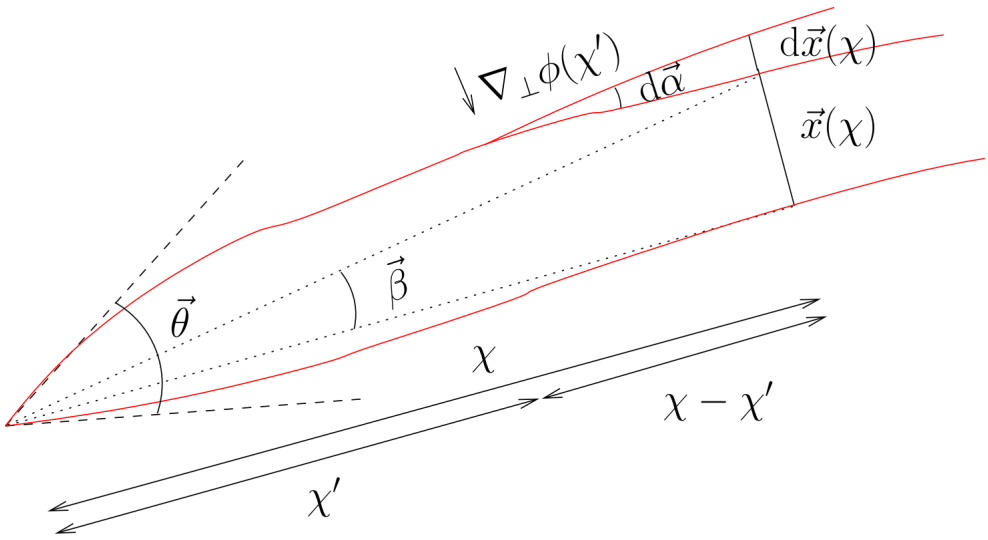


Figure 2.3: Illustration of the angular image distortion (from β to θ) at comoving distance χ that takes place due to a potential gravitational field (Kilbinger, 2014).

be found by removing the deflected contribution

$$\begin{aligned} x(\chi) &= x_0 - x_{\text{deflected}} \\ &= \chi\theta - \frac{2}{c^2}\int_0^\chi d\chi'(\chi-\chi')[\nabla_\perp\Phi(x,\chi') - \nabla_\perp\Phi^{(0)}(\chi)], \end{aligned} \tag{2.23}$$

where $\Phi^{(0)}(\chi)$ takes into account that the second ray path will also be deflected, thus contributing to the overall separation. The corresponding angle to the true separation would be $\beta = x(\chi)/\chi = \theta - \delta$ (known as the lens equation) leading to an angle of deflection not unlike equation 2.19

$$\delta = \frac{2}{c^2}\int_0^\chi d\chi' \frac{\chi-\chi'}{\chi}[\nabla_\perp\Phi(x,\chi') - \nabla_\perp\Phi^{(0)}(\chi)]. \tag{2.24}$$

Once again the born approximation can be used so instead of continually calculating x in $\nabla_\perp\Phi(x,\chi')$ we presume the observed light ray took a straight line path so that $x \approx x_0 = \chi\theta$. With this in mind, one can create a Jacobian matrix $\boldsymbol{A} = \partial\beta/\partial\theta$ so that one can map from image coordinates θ to source coordinates β,

$$A_{ij} = \delta_{ij} - \frac{2}{c^2}\int_0^\chi d\chi' \frac{(\chi-\chi')\chi'}{\chi}\frac{\partial^2}{\partial x_i \partial x_j}\Phi(\chi'\theta,\chi'), \tag{2.25}$$

where δ_{ij} is the delta function. Just in the same way as the deflection angle in equation 2.19 uses the gradient of the gravitational potential, one can rewrite the Jacobian also to be dependent on the gradient of the lensing potential ψ,

$$\begin{aligned} A_{ij} &= \delta_{ij} - \partial_i \partial_j \psi \\ \text{where } \psi(\theta,\chi) &= \frac{2}{c^2}\int_0^{\chi} d\chi' \frac{(\chi-\chi')}{\chi\chi'}\Phi(\chi'\theta,\chi'). \end{aligned} \tag{2.26}$$

An obvious effect of structuring the Jacobian with the use of a delta function is that in the case of no mass deflecting the photons path (i.e. $\psi = 0$) the Jacobian simply becomes the identity matrix $\mathbf{A} = \mathbf{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ which means the image coordinates are equivalent to the source coordinates $\beta = \mathbf{A}\theta = \mathbf{I}\theta = \theta$. The Jacobean in this form allows us to define the second derivatives of the potential as the scalar convergence κ and 2-component spin-2 shear $\gamma = (\gamma_1, \gamma_2)$ to introduce parameters that have direct relationships with the size and shape of the images we collect of distant galaxies,

$$\mathbf{A} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{2.27}$$

where

$$\kappa = \frac{1}{2}(\partial_1\partial_1 + \partial_2\partial_2)\psi = \frac{1}{2}\nabla^2\psi \quad \gamma_1 = \frac{1}{2}(\partial_1\partial_1 - \partial_2\partial_2)\psi \quad \gamma_2 = \partial_1\partial_2\psi. \tag{2.28}$$

The convergence and shear (κ, γ) parameters have not been chosen randomly but are direct consequences of requiring an intuitive understanding of the Jacobian. One can begin by splitting $\mathbf{A}$ between the parts that do not change the image equally in all directions (the anisotropic part) and the part that does (isotropic), which is given by the trace of the matrix,

$$tr(\mathbf{A}) = \sum_{i=1}^{2} \delta_{ii} - \partial_i\partial_i\psi = 2 - \nabla^2\psi = 2(1-\kappa). \tag{2.29}$$

Separating the trace reveals how the shear holds the anisotropic effects,

$$\mathbf{A} = \kappa\mathbf{I} - \Gamma, \tag{2.30}$$

$$\Gamma = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix}. \tag{2.31}$$

The isotropic nature of the convergence means it essentially acts as a magnifier changing the image's size but not its overall shape. We can see this by inverting the $\mathbf{A}$ as this reveals how

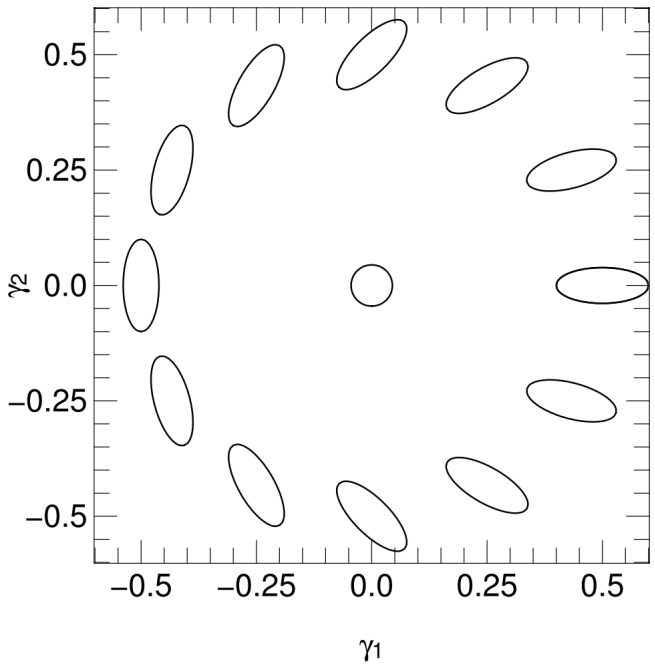


Figure 2.4: The level of elliptical distortion that can take place on a spherical 2-d image given different amounts of γ across its 2 components (Kilbinger, 2014).

the truth becomes the observed image ($\theta = \mathbf{A}^{-1}\beta$),

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\begin{pmatrix} 1-\kappa+\gamma_1 & \gamma_2 \\ \gamma_2 & 1-\kappa-\gamma_1 \end{pmatrix}. \tag{2.32}$$

So the prefactor $1/\det(\mathbf{A})$ acts as the magnifier multiplying the image coordinates. In our case that magnifier is $\mu = 1/\det(\mathbf{A}) = 1/((1-\kappa)^2-\gamma^2) \approx 1+2\kappa$ as κ and γ are both close to zero. The magnification is completely convergence dependent. On the other hand, the anisotropic element of lensing, the shear γ, can be summarised as not changing the size of the lensed object image but its shape. Consider a circular source which has been lensed. The lensing process will lead to the observed image being an ellipse. One can characterise the ellipticity ε (a combination of the semi-major and semi-minor axis a, b) through the inverse Jacobian's eigenvalues $\lambda_\pm$,

$$\varepsilon = \frac{a-b}{a+b} = \frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-} = \frac{|\gamma|}{1-\kappa}, \tag{2.33}$$

where $\lambda_\pm = (1-\kappa \pm |\gamma|)/(\det(\mathbf{A}^{-1}))$. The effect of shear on the image's shape can be seen clearly in Figure 2.4 where changes in γ_1, γ_2 make the circle more elliptical but also change in orientation.

2.3 Observables

2.3.1 Convergence

As stated above, gravitational lensing's observable effect is that images of galaxies we collect will be distorted in size and shape. However, for this phenomenon to be a cosmological probe, as hinted at in Chapter 1, one must be able to connect these observations to cosmological parameters of interest. Thankfully this is possible. The presence of matter courses gravitational lensing. Thus, the amount of distortion in our images gives us a gauge of how much matter was between the source and the observer on the photon's path. Consider the convergence equation 2.28 but one replaces the lensing potential with the integral from equation 2.26,

$$\kappa = \frac{1}{c^2}\int_0^{\chi'} d\chi \frac{\chi(\chi'-\chi)}{\chi'}\nabla^2\Phi, \tag{2.34}$$

where the integrand is magnified by a factor of χ^2, due to the laplacian in equation 2.28 being 2D (as it the referring to the image), whereas the in equation 2.26 the potential is 3D (as it can be anywhere in the universe). We can now use the Poisson equation $\nabla^2\Phi = 4\pi G a^2\rho$ where ρ is the matter distribution and rewriting in terms of perturbations to have a relation between convergence and cosmological parameters,

$$\begin{aligned}\kappa &= \frac{4\pi G}{c^2}\int_0^{\chi'} d\chi \frac{\chi(\chi'-\chi)}{\chi'}a^2\rho \\ &= \frac{3H_0^2\Omega_M}{2c^2}\int_0^{\chi'} d\chi \frac{\chi(\chi'-\chi)}{a\chi'}\delta,\end{aligned} \tag{2.35}$$

where the relation between mass density ρ and perturbation density δ is $\rho = \bar{\rho}\delta = (3H_0\Omega_m/a^3 8\pi G)\delta$. One obvious probe will be that if one knows the convergence field, one can infer the corresponding perturbation field. This is the crux of weak lensing surveys, requiring many galaxy images to understand the distributions between source and image. The issue becomes the fact that one cannot observe convergence directly. One may want to use an estimator such as the magnification μ, but the estimator can also be defined as the ratio of the image to source size $\mu = \theta/\beta$ one still requires information of the source area that is unavailable.

This does not mean magnification is completely undetectable. By considering Liouville's theorem, one can analytically prove the conservation of surface brightness Lieu (2004) states that "The surface brightness of an image of a lensed source coincides with the surface brightness of a source in the absence of lensing" (Petters et al., 2001). This means that not only does the size of the lensed image increase but also its flux. Once again, without

knowledge of the sources of intrinsic flux, one cannot gain information individually; however, one can detect it by looking at the distribution of galaxy fluxes across a large area. The 2 increases (image size and image flux) have conflicting effects in the context of a survey detecting millions of galaxies. The size increases mean the overall source density in any region will be decreased (due to the stretching of the image decreasing source count density), whereas the flux increase means sources will be detected that would otherwise be undetectable due to the flux limit of the survey (sources below the flux limit being to faint to be observed) (Hildebrandt et al., 2009). Depending on the effect, one will detect negative or positive angular cross-correlations (i.e. in the case of the flux limit, one will find many more than expected faint galaxies in a single region due to that region being lensed). These cross-correlations lead to a signal, the first detected using optically selected quasars from the Sloan Digital Sky Survey (SDSS). (Scranton et al., 2005). The signal-to-noise ratio one can expect from magnification can be very low (Van Waerbeke, 2010) which, when combined with the fact that flux correlations are intrinsic to survey distributions, makes the collection of cosmological information a difficult task (Heavens et al., 2013).

2.3.2 Shear

A more widely available observational and much higher signal-to-noise ratio comes from observing cosmic shear γ. In reality, one wishes to remove the magnitude part of the observation to have a shape-only based result, so the true observational is the reduced shear

$$g = \frac{\gamma}{1-\kappa}, \tag{2.36}$$

which from equation 2.33 shows that $|g| = \varepsilon$. Unfortunately, this is still an oversimplification. The ellipticity of equation 2.33 only describes shape not orientation, this issue can be mitigated by considering the two-component spin 2 shear to be a complex number $\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2i\phi}$ whereas ϕ passes from 0 to 2π the ellipsis rotates around π. Adding orientation means ellipticity becomes $\varepsilon = (a-b)e^{2i\phi}/(a+b)$ the only change this makes is one no longer detects the magnitude of the reduced shear but the full complex value. This would still be fine if every galaxy were inherently a perfect sphere of $\varepsilon = 0$ which is unfortunately not the case. All galaxies have an amount of source ellipticity ε^s which, when taken into account, gives a new image ellipticity (Seitz and Schneider, 1997)

$$\varepsilon = \frac{\varepsilon + g}{1 + g^*\varepsilon^s} \approx \varepsilon + g, \tag{2.37}$$

where the approximation is allowed under the weak-lensing regime of $\kappa, |\gamma|, |g| < 1$. Just as with magnification, one cannot gather information from the image without knowing the source. There is one major advantage which is that we have a much better understanding of intrinsic elliptical on a statistical level so that we can assume that the source shape and orientation of a galaxy to be completely random $< \varepsilon^s >= 0$ allowing the observed ellipticity to be an unbiased estimator of reduced shear

$$< \varepsilon >= g. \tag{2.38}$$

This also is not entirely the case as ellipticity is not completely random (see later sections) but small enough that the first shear signals could be detected on this assumption (Bacon et al., 2000).

Unfortunately, reduced shear estimation is not that simple. Up to now, we have presumed one has perfect knowledge of the galaxy images' ellipticity and that any distortion in that image (compared to the source) is exclusively due to weak lensing. In both cases, the assumption is incorrect, and if one does not mitigate for these assumptions, heavy biases would be introduced (Massey et al., 2011). Starting with the later assumption that distortions are weak lensing exclusive. This is not the case. Images of galaxies one receives from telescopes are blurrier than they should be. This blurring can course, add smoothing and distortion to the image and is known as the point spread function (PSF) of the observations. There are multiple sources of the PSF (Kitching et al., 2011) :

- the way the individual telescope is set up, from detected photons to pixellated image, will introduce a unique PSF which changes with image position
- photons detected by ground telescopes will be refracted due to atmospheric conditions
- any slight movement by the telescope while observing can lead to additional smoothing (as the slightly moving image would be averaged out). Multiple effects can course such movements as vibrations and wind.
- additional noise due to the accidental detection of background photons, the noise is usually modelled as gaussian for ground telescopes and Poisson for space-based telescopes (due to the low number of background photons)

There are 2 forms of bias that can occur if PSF effects are not dealt with correctly; Firstly, the smoothing of the image reduces its ellipticity; thus, a multiplicative bias is introduced that reduces overall observed shear. Secondly, any anisotropies in the PSF will invite additive errors to the ellipticity observed (Mandelbaum, 2017).

To remove the PSF before, one must know what the PSF for that observation is (as atmospheric effects such as weather and temperature are time-dependent, one cannot have a constant PSF even for the same telescope). Gathering said knowledge is usually achieved by observing stars close to the galaxy. As far away stars can be considered point sources of which their actual spherical shape is known, any distortion in shape can be attributed to (and thus used to model) the PSF. One then needs to interpolate the PSF to other positions in the sky so that it can be removed from nearby galaxies being observed. There are varying methods to achieve the correction interpolation, the most common of which is an empirical approach of fitting a low-order spacial polynomial function to the stars and then smoothly changing variables of these polynomials depending on where your galaxy is in the sky. Such an approach was considered sufficient for space-based telescopes where the PSF was instrumental in origin so well modelled by smoothly varying functions (Van Waerbeke et al., 2002). However, future wide-field surveys by ground-based telescopes with short exposure times such as the LSST have much more significant atmospheric effects that do not follow smooth functions requiring more sophisticated fitting methods (Chang et al., 2012). Alternatively, one can use a physics-based forward modelling approach in which one directly simulates the PSF as a function of position by ray tracing the photons through physical models of the telescope and then choosing the appropriate model by which one simulates nearby stars' PSFs best. This reduces the interpolation issue as it only needs to interpolate the galaxy's coordinates on the fine space grid of the model as opposed to interpolating between the stars themselves. A method such as this entirely depends on the simulated model being correct, so mistakes in physics have detrimental effects. This would be particularly important when trying to correctly forward model chaotic events such as atmospheric effects, so with that in mind, this approach works using space-based telescopes and has been successful in the past with data such as the COSMOS survey (Leauthaud et al., 2007) (Rhodes et al., 2007).

One should not limit the effects of the point spread function to the size and shape of the observed object (size and shape being second moments in this case). Before 2020 the formalism used to quantify PSF error (and thus work out the effects it had on weak lensing shear bias) did not consider higher moments and so would predict 0 biases as long as size and shape work correctly found regardless of higher moments (Jarvis et al., 2016). Unfortunately, this is not the case, as additional additive and multiplicative biases for cosmic shear were found for Euclid that went beyond those predicted by their second moment formalism with a realistic telescope PSF (Schmitz et al., 2020). The most probable suggestion for the added bias is in higher-order errors. To investigate its effects Zhang et al. (2022) measured the fourth radial moment (kurtosis) of simulated images and related them to their PSF models and

shear measurements to find that errors in fourth moments alone can contribute to $\sim 0.05\%$ of shear bias, negligible for current surveys but essential going forward for LSST and EUCLID.

Another PSF-based effect that needs to be correctly considered is the Chromatic PSFs. What this refers to is the fact that all modelling techniques above have assumed that the PSF is only a function of location (and even then, it slowly changes) and that each galaxy has one PSF. However, the PSFs are wavelength-dependent as well. This can be due to multiple reasons, the diffraction limit of the telescope lens, refractions in the atmosphere, which have a wavelength-dependent trajectory and sensor contributions within the telescope (Meyers and Burchat, 2015). This would affect the PSF modelling methods mentioned above as they all use individual stars (either as a point of model creation or verification) that release different light (wavelength-wise) compared to the galaxies to which we apply the model (Cypriano et al., 2010). This wavelength dependence can also break the second assumption as galaxies contain colour gradients, emitting photons of multiple wavelengths, potentially multiple PSFs (Er et al., 2017). With that said current studies suggest that through image simulation in different bands, one can accurately model the bias caused by changes in wavelength and correct it in the final result, even in time-dependent atmospheric conditions seen in LSST ground-based surveys (Carlsten et al., 2018).

The location-only dependency of the PSF continues to break down if one considers the effects and imperfections in the detectors themselves. For example, weak lensing surveys usually observe faint sources, while the models above use bright starts, so any flux dependence in the actual PSF would lead to inaccuracies in our models. One such flux-dependence is the 'brighter-fatter' effect (Antilogus et al., 2014) in which brighter sources lead to charge building up within a pixel. This induces an electric field which deflects further light away from charged pixels to nearby pixels. The outcome of this deflection is that brighter objects can seem broader than faint objects. Thankfully this effect can be modelled by considering the correlations in nearby pixels (Coulton et al., 2017).

2.3.3 Ellipticity Estimation

Presuming one had modelled the PSFs correctly, considering all the effects mentioned above (location dependence, interpolation technique, forward model, wavelength dependence, flux dependence etc.). Outside of effects caused and biases created by the pixelation process and other systematic effects (see below), one can presume that the image being observed is an image that is exclusively distorted from weak lensing. As mentioned above, one would measure the ellipticity to gain a statistical shear. However, this is an idealised case. Not all galaxies follow perfect ellipses. This is doubly the case when the images observed are pixelated. This is a galaxy-by-galaxy process which can be computationally expensive. Thus,

many models have been created to complete the non-trivial task of correctly estimating shears.

A popular method is weighted quadrupole moment (Okura and Futamase, 2012). Consider an image consisting of a smooth light profile with fluxes that differ as a function of position $I(\theta)$ where $\theta = (x, y)$ the centre of the object can be defined as the first moment

$$\mu_i = \frac{\int I(\theta)\theta_i d^2\theta}{\int I(\theta) d^2\theta}. \tag{2.39}$$

The weighted second moment measures the covariance of luminosity around a central point,

$$Q_{ij} = \frac{\int (\theta_i - \mu_i)(\theta_j - \mu_j) W(\theta - \mu) I(\theta)}{\int W(\theta - \mu) I(\theta) d^2\theta}, \tag{2.40}$$

where $W(\theta - \mu)$ is the weighting function usually chosen to maximise the signal-to-noise ratio (SNR) and to minimise the contribution of far-away noise pixels. Using these secondary moments, one can make shape estimates for the ellipticity such as distortion,

$$\varepsilon = \frac{Q_{xx} - Q_{yy} + 2iQ_{xy}}{Q_{xx} + Q_{yy} + \sqrt{Q_{xx}Q_{yy} - 2Q_{xy}^2}}. \tag{2.41}$$

Such estimates for shear have been successful in the past with choosing the correct weighting system leading to the correct shear values being recovered with a 10% systematic uncertainty (Bacon et al., 2001) other advantages include the fact that quadrupoles do not presume a light profile for the galaxy so do not apply a model bias. The lack of model bias means unbiased estimates for the galaxy shear can even be found for galaxies with complex morphology (Bernstein, 2010). However, this argument only holds if we negate the fact that the method is extremely weighting-dependent where the weightings themself over their own model bias.

An alternative method is based on a decomposition of the galaxy image into shapelets. Once again, consider the flux profile of the image $I(\theta)$. It can be decomposed into a set of basis functions

$$I(\theta) = \sum_n f_n B_{\mathbf{n}}(\theta), \tag{2.42}$$

where $\mathbf{n} = (n_1, n_2)$, $f_{\mathbf{n}}$ are the basis coefficients and $B_{\mathbf{n}}$ is the 2-dimensional basis function. This can also be written in terms of 1-dimensional Basis functions

$$B_{\mathbf{n}} = \prod_{i=1}^{2} B_{n_i}(\theta_i), \tag{2.43}$$

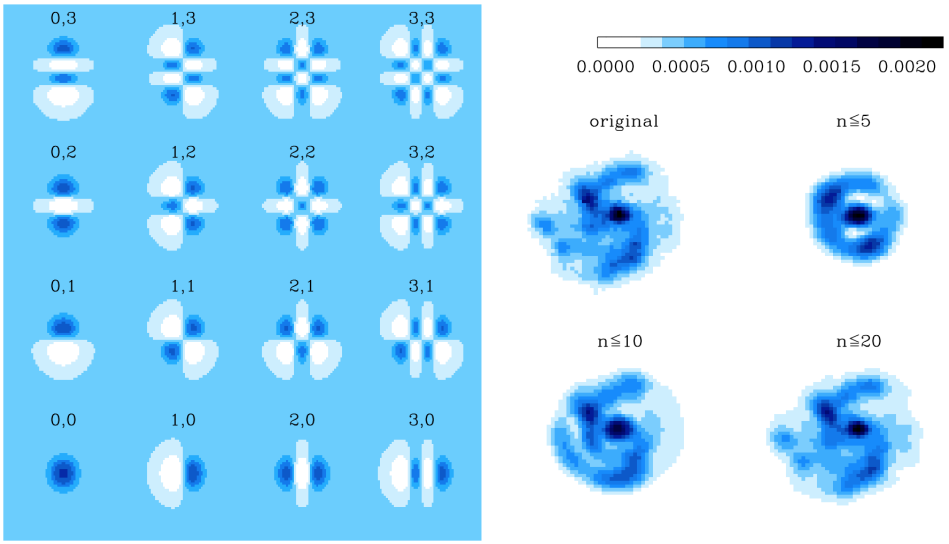


Figure 2.5: *left* : The different shapelet components made through different combinations of (l,m). *right* : How using different numbers of these components can replicate an original image with reasonable parity already possible with $n \leq 20$ (Refregier, 2001).

$$B_{n_i}(\theta_i) = [2^{n_i}\pi^{0.5}n_i!\beta]^{-0.5}H_{n_i}(\frac{\theta_i}{\beta})e^{-2\frac{\theta_i^2}{2\beta^2}}, \tag{2.44}$$

where β is a scale usually around the radius of the image while H_n is the Hermite polynomial of order n. The coefficients $f_{\mathbf{n}}$ are found using

$$\int_{-\infty}^{\infty} d^2\theta f(\theta)B_{\mathbf{n}}(\theta). \tag{2.45}$$

As the basis functions form an orthogonal set. This effect can be seen clearly in Figure 2.5. Figure 2.5.a shows 2 different 2-dimensional shapes each basis function represents (where the dark and light regions correspond to positive or negative values) when applied to an image (Figure 2.5.b). In this case, an HST image, using more basis functions, improves the image to the point where $n < 20$ is almost identical to the original (Refregier, 2001). Shapelets can be seen as a generalisation of the quadrupoles to all available multi-pole methods as the shapelets when $n_1 + n_2 = 0$ are equivalent to Gaussian weighted quadrupoles (Refregier and Bacon, 2001).

The shapelets break down an image into a set of shape components, each of which measures the effect of shear would be easier than the combined complicated image. Consider our $I'(x)$ represents the unsheared flux image which can be represented in a first-order approximation

$$I(x) \approx (1 + \gamma_i S_i)I'(x), \tag{2.46}$$

where γ_i and S_i are the shear and shear operators, respectively. If our sheared image is decomposed into basis functions as seen above, we can express a relationship between the lensed and unlensed basis coefficients (f_n, f'_m)

$$f_n = (\delta_{nm} + \gamma_i S_{inm}) f'_m, \tag{2.47}$$

where S_{inm} is the shear matrix. Using the fact that the average shapelet coefficient $< f_n >$ before lensing must be rationally invariant and that shear breaks this symmetry by introducing a preferred direction, one can construct a measure of this symmetry breaking for each shear coefficient

$$\tilde{\gamma}_{1n} = \frac{f_n - < f_n >}{S_{1nm} < f_m >}, \tag{2.48}$$

when n_1 and n_2 are even and

$$\tilde{\gamma}_{2n} = \frac{f_n - < f_n >}{S_{2nm} < f_m >}, \tag{2.49}$$

when n_1 and n_2 are odd.

With the individual component shear, estimated one can combine them for a full shear estimation

$$\tilde{\gamma}_i = \frac{\sum_n w_{in} \tilde{\gamma}_{in}}{\sum w_{in}}, \tag{2.50}$$

where w_{1n} and w_{2n} are the weights which are set to zero when n_1 and n_2 are odd-odd or even-even respectively. When not set to zero, the weights are optimised to minimise estimator component variance. Such an estimator is unbiased and linear (Refregier, 2001) furthermore, being in this shapelet form allows for multiple advantages. It turns shear into simple matrix calculations. It is linear in galaxy intensity, which avoids some forms of bias. It uses all available shape information, whereas the quadrupole method above only uses 2-order shapelets which can introduce unstable and ill-defined quantities (Kaiser et al., 2000). There are, however, a few criticisms of the shapelet approach. For both practical reasons and not unintentionally overfitting the noise, one is forced to use a truncated set of basis functions, so the full shapelet expansion is not used to describe the galaxy. This leads to the shapelet-constructed images resembling gaussian intensity profiles while galaxies follow Sersic profiles (Sérsic et al., 1963) leading to underestimates in some quantities. Studies have taken place to quantify how limiting the shapelet method (Melchior et al., 2010) is in real-life scenarios. The two realistic conditions chosen are those that have already been mentioned; a limited number of sample modes and intrinsic galaxy shapes that are not restricted to shapelet models. By simulating galaxy shapes based on a Sersic profile, the study found that highly elliptical galaxy shapes cannot be accurately modelled and are biased towards less elongated galaxies. Issues compound when higher Sersic indices are included in the profile (see below),

at which point shear estimates are biased due to the preference for round Gaussian models; thus, the use of shapelets for practical shear estimation has been heavily questioned.

A more direct alternative route is to use model fitting. As with the PSF approaches, we fit the galaxy images to the parametric model. We most often presume the Sersic profiles mentioned above, which take the following form (Kacprzak et al., 2014),

$$I(x) = A \exp(-k[(x-x_0)^T C^{-1}(x-x_0)]^{0.5n}), \tag{2.51}$$

where A is the amplitude, x_0 is the centre of the observed image, C is the covariance matrix and $k = 1.9992 - 0.3271n$ where n is the Sersic index profile. The profile, although simplistic, can be compelling with Sersic indices of $n = 1$ (exponential profile) and $n = 4$ (de Vaucouleurs profile) being commonly used models for the disk and bulge of observed galaxies, respectively. Numerical optimisation algorithms are usually used (such as steepest descent and adaptive grid search) to find the parameters that best suit the image, and then using the ellipticity is defined for this model. The method can be efficient; however obvious pitfalls come from model fitting methods; systematic effects can occur through the choice of optimisation algorithm, and model bias can occur if the model oversimplifies the galaxy shape (studies through simulation have shown that if a galaxy is comprised of 2 particular shapes such as a bulge and a disk, then biases are introduced that need to be taken into account (Voigt and Bridle, 2010)).

There also exist methods that do not use the shears of individual galaxies at all and thus are entirely independent of galaxy shape. One such method is the stacking method. This method is based on the presumption that intrinsic galaxy shapes are uncorrelated (an assumption that will be scrutinised in more detail below) so that if you were to unlensed images of the galaxy on top of each in pixel space, the resulting image should be a circle thus the ellipticity of the resulting image would be due to exclusively to the shear and provide an adequate estimation. Furthermore, this method avoids common biases of the methods above; the sum of independent samples should have a close to Gaussian distribution due to the central limit theorem (Lewis, 2009), so the statistics of the stacked images are known without knowing the distribution of individual galaxies, those also removes model bias as were not presuming the galaxies have a particular shape. Alternatively, an improved stacking method is known as spectral stacking (Hosseini and Bethge, 2009) where one completes the same procedure but in Fourier Space using the image's power spectrum. The approach's success can be seen in the GREAT08 challenge (Bridle et al., 2009). In this 6-month competition, participants were asked to analyse 30 million simulated galaxies with different shapes, signal-to-noise rates, point spread functions and ellipticities to estimate their shear. From this challenge,

the spectral stacking approach won the competition with a score of 210.9 (compared to the image stackings 131.4).

2.4 Shear Prediction

2.4.1 Shear Power Spectra

Under the assumption that the methodology for taking noisy pixellated images of galaxies and extracting shear estimates is well understood and biases limited, this alone does not allow science to take place. The shears of individual galaxies, although observable, are not currently predicted by the standard model. As one can see from equation 2.35 if one wishes to predict the lensing effects along a line of sight (presumable between the observer and a distant galaxy), one requires accurate knowledge of the matter distributions (both dark and baryon) along that line of sight. The knowledge that is not currently available. What can be measured and predicted, though, are statistical lensing effects such as the degree to which lensing quantities (such as shear convergence and lensing potential) correlate with each other this is because the statistical quantities of matter (through the correlation function and power spectrum) can be predicted through the standard model as seen in Figure 1.10. As the cosmological principle is held for the matter distribution, so does it hold for the convergence/shear distribution meaning that the 2-point correlation function of these lensing quantities is invariant under translation and rotation.

Consider the 2-point correlation function of the convergence $< \kappa(\omega)\kappa(\omega') >$ as seen above. This is the assembled average over 2 angular positions separated by ω'. When expressed in Fourier space, one can define the convergence power spectrum P_κ,

$$\langle \tilde{\kappa}(\ell)\tilde{\kappa}^*(\ell')\rangle = (2\pi)^2\delta_D(\ell-\ell')P_\kappa(\ell), \tag{2.52}$$

where δ_D is the Dirac delta function, $\tilde{\kappa}, \tilde{\kappa}^*$ is the Fourier transforms of the convergence, and its complex conjugate, respectively, and ℓ is the Fourier conjugate of θ, a 2D wave vector (which differs slightly from equation 1.77 due to us using angular position). As seen previously, the convergence power spectrum depends only on ℓ because of statistical homogeneity and isotropy. We can now square our expression from 2.35, and convert it to Fourier space to directly relate the power spectrum of the convergence to the power spectrum of the density contrast P_δ,

$$P_\kappa(\ell) = \frac{9}{4}\Omega_m^2(\frac{H_0}{c})^4\int_0^\chi d\chi(\frac{\chi(\chi'-\chi)}{a\chi})^2P_\delta(k=\frac{\ell}{\chi},\chi). \tag{2.53}$$

Through equation, 2.53 were in a position where one can make actual cosmology. We observe and compare the convergence power spectrum to the predicted convergence power spectrum given to us through the standard model's predicted density contrast power spectrum. With observation and a prediction, one can infer the parameters of the standard model that appear in equation 2.53 such as matter density Ω_m. It is worth noting that to get to equation 2.53 many approximations were made that will be considered in more detail below; however, there is a pressing matter of the fact that the above has been referring to the convergence power spectrum but as we stated before were observing the shear not the convergence through that method mentioned above. Thankfully we can relate one to the other. In fact, by evaluating the distortion matrix of equation 2.32 for the case of 2 masses (an acceptable approximation under the thin-lens approximation), one can derive an expression for shear in terms of convergence through the trace-free part.

$$\gamma = \pi^{-1} \int d^2\theta' \kappa(\theta') D(\theta - \theta'), \tag{2.54}$$

where $\theta = (x, y)$ in Cartesian coordinates,and $D(\theta) = (y^2 - x^2 - 2ixy)/|\theta|^4$ is a complex quantity. Shear and convergence are so closely related that one can recover the convergence from shear measurements through Kaiser Squares Inversion (Kaiser et al., 1993). This method is based on the fact that equation 2.54 is essentially a convolution (where a convolution is defined as an integral that expresses the amount of overlap as one function g passes over another function f in t-space as the following $[f * g](t) = \int_0^t f(\tau) g(t - \tau) d\tau$ where in out case g and f are replaced with κ and D in θ-space instead of t-space) and that in Fourier space a convolution becomes a product allowing for a simpler relation

$$\tilde{\gamma} = \frac{1}{\pi} \tilde{\kappa}(\ell) \tilde{D}(\ell), \tag{2.55}$$

where $\ell = (\ell_1, \ell_2)$ and $\tilde{D}$ is the Fourier transform of the filter form equation 2.54 defined as $\tilde{D}(\ell) = (\ell_1^2 - \ell_2^2 + 2i\ell_1\ell_2)/\ell^2$ so one can simply take the inverse of the integrand to get from one to the other (shear to convergence). Even more, use full that if we move to polar coordinates and drop the constant term, the relation simplifies more to

$$\tilde{\gamma}(\ell) = \frac{(\ell_1 + i\ell_2)^2}{\ell^2} \tilde{\kappa}(\ell) = e^{2i\beta} \tilde{\kappa}(\ell), \tag{2.56}$$

where β is the polar angle of the wave vector ℓ. We can now apply this relation to our actual observable, the 2-point-correlation function of the shear and Fourier transform it just as we

did with the convergence in equation 2.52 as

$$\langle\tilde{\gamma}(\ell)\tilde{\gamma}^*(\ell')\rangle = \langle e^{2i\beta}\tilde{\kappa}(\ell)e^{-2i\beta}\tilde{\kappa}^*(\ell')\rangle = \langle\tilde{\kappa}(\ell)\tilde{\kappa}^*(\ell')\rangle = (2\pi)^2\delta_D(\ell-\ell')P_\kappa(\ell). \quad (2.57)$$

Thus it is clear that the power spectrum of the convergence and the shear are equivalent ($P_\gamma = P_\kappa$), so as long as one observes the shear power spectrum, one can infer the matter density contrast power spectrum.

With this equivalence known, we can now consider the shear 2-point correlation function as our observable of choice. We will consider shear 2-point statistics as it noted by Bartelmann and Schneider (1999), by considering pairs of $(\theta, \theta+\theta')$ we can rewrite our shear quantity in terms tangential component $\gamma_t = -Re[\gamma e^{-2i\phi}$ and its cross-component $\gamma_x = -Im[-2i\phi]$ where ϕ is the polar angle of the separation vector θ.By convention, one adds a minus sign so that γ_t is positive for the tangential alignment around an overdensity while a positive cross-component shear has a $\pi/4$ rotation compared to the tangential shear. On the other hand, the tangential shear component is negative for a radial alignment around an under-density. Under this 2-component format, the 2-point-correlation function is no longer a single function but 3 based on the statistical interaction of the components $(\langle\gamma_t,\gamma_t\rangle, \langle\gamma_x,\gamma_x\rangle, \langle\gamma_t,\gamma_x\rangle)$, the last of these vanishes to zero in a parity-symmetric universe. The other 2 are linearly combined to give a correlation function $\xi_\pm$, which relates to the convergence power spectrum as

$$\xi_\pm = \langle\gamma_t,\gamma_t\rangle \pm \langle\gamma_x,\gamma_x\rangle = \frac{1}{2\pi}\int_0^\infty d\ell\ell P_\kappa(\ell)J_{0,4}(\ell\theta), \quad (2.58)$$

where $J_{0,4}$ refers to the 2 Bessel functions (J_0, J_4) that correspond with the 2 correlation function operators $+,-$ respectively. The 2 different correlation functions are essentially linearly filtered versions of the convergence power spectrum. The filters in question are the 2 Bessel functions that work differently at small $\ell\theta$. For example, J_0 remains almost constant for $\ell\theta \ll 1$ while J_0 behaves as $(\ell\theta)^4$ so ξ_- is a lot more localised version of the convergence power spectrum. Even with this filtering process, information is not lost as the full convergence power spectrum can be recovered from just one of these correlation functions,

$$P_\kappa(\ell) = 2\pi\int_0^\infty d\theta\theta\xi_+J_0(\ell\theta) = 2\pi\int_0^\infty d\theta\theta\xi_-J_4(\ell\theta). \quad (2.59)$$

For complete consistency, one can also recover one correlation function from the other though recovering $\xi_-(\theta)$ from $\xi_+(\theta)$ is the desirable option as it only requires an integral of $[0,\theta]$ as opposed to the much less practical solution of the reverse ($\xi_+(\theta)$ from $\xi_-(\theta)$) which requires an infinite range of integration.

Unfortunately, evaluating the convergence power spectrum straight from the correlation functions above is incredibly in-practical. Consider equation 2.59 to fully recover the convergence power spectrum one would come across 2 issues. Firstly we do not have perfect knowledge of $\xi_{\pm}(\theta)$, every galaxy observation is noisy, and many shear estimations have to be made, as demonstrated above. Furthermore, even if we did have a method to perfectly observe the shear of distant galaxies, the integral of equation 2.59 requires knowledge of correlation function at an angular separation of $\theta = \infty$ knowledge unreachable to us as it requires observations outside of our light cone. With this in mind, we use an estimator (Schneider et al., 2002) to estimate the correlation function based on galaxy observations,

$$\hat{\xi}_{\pm} = \frac{\sum_{ij} w_i w_j (\varepsilon_{t,i}\varepsilon_{t,j} \pm \varepsilon_{x,i}\varepsilon_{x,j})}{\sum_{ij} w_i w_j}. \tag{2.60}$$

The sums take place over a pair of galaxies (i, j) at an angular separation of $|\theta_i - \theta_j|$ (so that their angular positions are θ_i, θ_j respectively). For each galaxy, one observes its ellipticity ε and derives its elliptical tangential and cross components $\varepsilon_t, \varepsilon_x$. We also assign a weight w to each galactic observation. By these means, we can down-weight noisy observations by assigning small values of w. Presuming that the observable ellipticity of a galaxy image at angular position θ_i takes a form of $\varepsilon_i = \varepsilon_i^s + \gamma(\theta_i)$ where ε_i^s is the intrinsic ellipticity which is valid under the weak lensing assumption $|\gamma| \ll 1$.

To evaluate the effectiveness of this estimator, we must consider the expectation value of the estimator $\langle\hat{\xi}\rangle$ as we will be taking the assembled average. if we first evaluate the assembled averages of the actual cross and tangential ellipticities, we get the following term,

$$\langle\varepsilon_{t,i}\varepsilon_{t,j} + \varepsilon_{x,i}\varepsilon_{x,j}\rangle = \sigma_{\varepsilon}^2 + \xi(\theta), \tag{2.61}$$

where σ_{ε}^2 is the dispersion of the intrinsic galaxy ellipticity, as this is the case by taking the ensemble average of the estimator, we gather the original correlation function, showing it to be an unbiased estimator $\langle\hat{\xi}_{\pm}\rangle = \xi_{\pm}$. Other than solving the issues mentioned above, there are other advantages of the estimator, such as the fact that it has an easily calculable covariance matrix, but most importantly, no knowledge of the mask geometry is required, just what galaxies are within the mask. This key advantage makes it the origin of other 2-point correlation function estimators designed to give a higher signal-to-noise ratio in specific angular scales and reduce correlations on different angular scales (a consequence of the original estimator.

Such a derived estimator includes the top hat shear dispersion (Kaiser et al., 1992) which takes the average root mean square of the shear for an aperture of radius θ (ie.$\langle|\gamma|^2\rangle(\theta)$).

Just as with equation 2.58 the estimator can be reexamined as a filter on the convergence power spectrum,

$$\langle|\gamma|^2\rangle(\theta) = \frac{1}{4\pi}\int_0^\infty d\ell\ell P_\kappa(\ell)\frac{2[J_1(\ell\theta)]^2}{(\ell\theta)^2}, \tag{2.62}$$

J_1 as a filter function is a low-pass filter on the power spectrum, this means that just like J_0 the filter is constant for $(\ell\theta) \ll 1$ but unlike J_0 which oscillates for large $(\ell\theta)$ with the amplitude slowly decreasing at $(\ell\theta)^{-1/2}$ the J_1 filter decreases steeply as $(\ell\theta)^{-3}$ at large values limiting its use. Thus estimator is light on usable information but returns very high signal-to-noise ratios for any given θ thus was a strong candidate for early cosmological studies where the signal-to-noise ratios were lower than are technologically available now. More modern studies have used the aperture mass dispersion (Schneider et al., 1998) as the statistic of choice. For this statistic, one defines a circular region with a characteristic scale θ and collects the mean tangential shear concerning the centre ϑ,

$$M_{ap}(\theta,\vartheta) = \int d^2\vartheta' Q_\theta(|\vartheta-\vartheta'|)\gamma_t(\vartheta'), \tag{2.63}$$

where Q_θ is the weighted filtered function. An alternative filter function can be derived to match the aperture mass to the convergence in both cases, and many approximate functions have been used in the past, including fourth-order polynomials and Gaussian functions. As with all previous estimators, it can be written in terms of a filter on the power spectrum.

$$M_{ap}(\theta,\vartheta) = \frac{1}{4\pi}\int_0^\infty d\ell\ell P_\kappa(\ell)\frac{576[J_4(\ell\theta)]^4}{(\ell\theta)^4}. \tag{2.64}$$

This is by far the most localised filter behaving as $(\ell\theta)^4$ and oscillatory decreasing as $(\ell\theta)^{-5}$ for small and large values, respectively. In both cases (top-hat shear and aperture mass), the statistics need estimators to be practical. These estimators use linear combinations of the original estimator mentioned above $\hat{\xi}_\pm$.

These seemingly inferior estimators offer advantages in easily separating the shear power spectrum into E and B modes. A brief return to the shear field is required to appreciate the significance. Fields can be defined in terms of their spin weights which explain how the field changes if an angle rotates the coordinate system θ (Heavens, 2008). If said rotation leads to a field change of $f \to fe^{si\theta}$, then s is the spin weight. As an example lensing potential ψ does not change under a coordinate rotation system thus $\psi = \psi ee^{si\theta}$, $e^{si\theta} = 1$ which means the lensing potential has a spin weight of $s = 0$ (this is more routinely known as a scalar field). The shear field, however, has a spin weight of 2 (i.e. it changes as $\gamma \to \gamma e^{2i\theta}$). This means that the shear field changes from its real to imaginary component (or pure γ_1 to γ_2

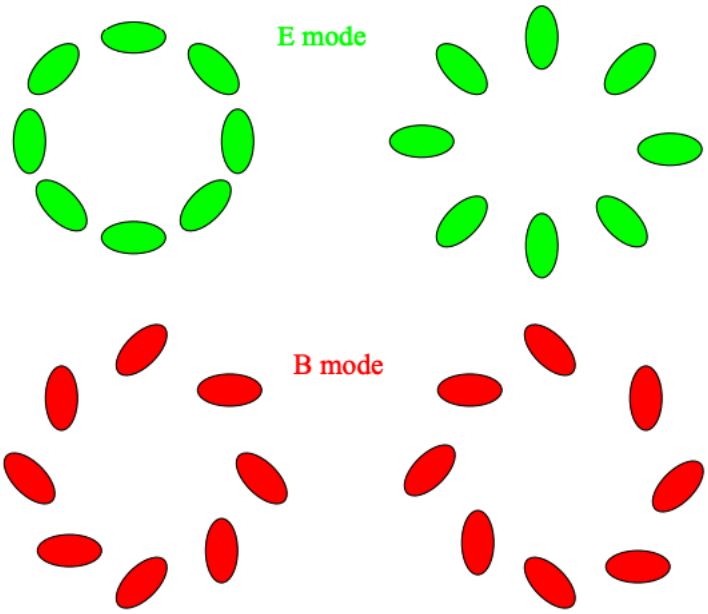


Figure 2.6: The different configurations of distorted elliptical images of galaxies that make the E-modes and B-modes of the lensing power spectrum (Van Waerbeke and Mellier, 2003).

every 45 degrees (alternatively, it returns to its original value every 90 degrees). The general quality of a spin 2 object is that it can be written as the second derivative of a complex field ϕ in modal form.

$$\gamma = \frac{1}{2}\partial\partial(\phi_E + i\phi_B), \tag{2.65}$$

where ϕ_E and ϕ_B are the modes of the complex field, these modes are defined by saying that if we have vector field u that is the gradient of the complex field (i.e. $u = \nabla\phi$), then the second derivatives of the E and B modes give the gradient and curl of that vector field respectively ($\nabla^2\phi_E = \nabla\cdot u, \nabla^2\phi_B = \nabla\times u$). In the case of shear, the complex field in question is the lensing potential, and due to this modal separation, the shear itself can also be separated into its 'E' and 'B' modes γ_E, γ_B. These shear modes have a physical interpretation as well as a mathematical description. For example, consider Figure 2.6 the green top images represent how spherical galaxy images would distort and align with just E-mode shear, whereas the bottom red images are the same but for B-mode shear. Furthermore, the images on the left represent how the modes distort images for central overdensities, whereas the right images are distortion caused by central under densities. Also, note that every E-mode galaxy image can be rotated 45 degrees to become its corresponding B-mode (thus the spin 2 nature of shear).

This modal structure can be continued into our chosen statistics for extracting cosmology from weak lensing observations. We can rewrite the 2-point correlation functions in terms of E and B modes,

$$\xi_{\pm}(\theta) = \frac{1}{2\pi}\int d\ell\ell J_{0,4}(\ell\theta)[P_{\kappa}^{E}(\ell) \pm P_{\kappa}^{B}(\ell)]. \tag{2.66}$$

This is a Hankel transform of the convergence power spectrum and can be easily and quickly integrated numerically (Hamilton, 2000). However, as one can see in equation 2.66 when expressed as standard 2-point correlation functions, the E and B modes are coupled and cannot be separated, which brings us back to the advantages of the derived 2-point statistics, which do have the ability to separate the 2 modes. The separated modal statistics can be shown given a derived 2-point statistic χ,

$$\chi_{E,B} = \frac{1}{2\pi}\int d\ell\ell P_{\kappa}^{E,B}(\ell)\tilde{U}^{2}(\ell), \tag{2.67}$$

where $\tilde{U}$ is a statistic-specific power spectrum filter. As with all statistics covered so far, estimators are required, which also have the mode separating qualities,

$$\hat{X}_{E,B} = \frac{1}{2}\sum \theta_i \Delta\theta_i [F_{+}(\theta_i)\xi_{+} \pm F_{-}(\theta_i)\xi_{-}(\theta_i)], \tag{2.68}$$

where $\Delta\theta$ is the bin width and F is the statistic-specific correlation function filter. The 2 filters are Hankel transforms pairs, so a relation exists that gets you from one to the other. For a complete list of filters and their corresponding 2-point statistic, please see the work of Kilbinger et al. (2013). Depending on the statistic chosen, it is easier to derive one filter and relate it to the other. For example, $\tilde{U}$ is just the Fourier transform of the real space filter function Q_{θ} in equation 2.63, and then from the relation, the correlator filter functions are derived.

It is now clear that the correlators offer this modal splitting advantage, but it is still not clear why this is an advantage in the first place. Consider again the shear modes expressed through equation 2.65, in that example, the complex field ψ to which shear is the second derivative is the lensing potential. This is particularly important as the lensing potential is a real field void of an imaginary component, so for the case of weak lensing, $\psi_E = \psi$ and $\psi_B = 0$. In other words, the shear field should not have a B mode. Thus the detection of a B mode (or, more superficially, the B model of the shear power spectrum) may indicate errors in the observation or analysis of the data, such as systematic errors in setup. Although not observable now due to only existing on the per cent level compared to E-modes, there are reasons for naturally occurring B-modes from weak lensing analysis. This includes B-mode correlation in the intrinsic shape of the galaxies, selection biases in the size and magnitude

of the observed galaxies and higher-order terms neglected by the approximations in place. Even with all the above errors, systematics, and physical phenomena accounted for, there would still be a degree of B-mode to a weak lensing survey, this is due to the universe not being isotropic on all scales, so any deviation from the local spacial isotropy assumed by standard cosmology would induce a B-mode in the shear (Pitrou et al., 2012). Unlike the B mode reason seen above the anisotropy, B-modes arise on all cosmological scales. This has at least 2 potential uses. Firstly, any bounds on a scale of B-modes detection can be used as a constraint on the isotropy of the universe (allowing further evidence for the cosmological principle). On the other hand, understanding local isotropy in the universe can help remove the B-modes in analysis, disentangling them from the non-cosmological B-modes.

2.4.2 Approximations

As mentioned above, the shear power spectrum as described in equation 2.53 can not be described in such a way without the need for approximations. The most physically changing is the flat-sky approximation. Up to this point, the coordinate system of choice has been Cartesian as in $\theta = x, y$ to say all observations are observed from a flat surface. This is not the case; observations are made by earth-bound telescopes that rotate with the earth or orbiting satellites. In both cases, the observations are on a curved surface which the normal power spectrum does not describe. Under this circumstance, instead of writing a function $f(x,y)$ as a summation of sine waves, we write the function in spherical coordinates, i.e. $f(\theta,\phi)$ were the coordinates of the choice point to locations on a surface of constant r. Under this system, the cartesian coordinates are related by $x = r\cos(\phi)\sin(\theta)$ and $y = r\sin(\phi)\sin(\theta)$ where ϕ is the azimuthal angle in the x-y plane from the x-axis where $0 \leq \phi \leq 2\pi$ and θ is the polar angle from the positive z-axis where $0 \leq \theta \leq \pi$.

Under this system, a normal Fourier transform to gather the power spectrum is no longer sufficient. Whereas with Fourier transforms, we use an infinite set of sin and cos waves to define a function. Now we use a set of spherical harmonics defined as

$$Y_{\ell m} = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos(\theta))e^{im\phi}, \tag{2.69}$$

where ℓ is the multipole representing a given angular scale of the sky to approximately π/ℓ in degrees and its values indices$\ell = 0,\ldots,\infty$. m is an index restricted by $-\ell \leq m \leq \ell$ and P_ℓ^m are the Legendre polynomials. With this in mind, we can take a sum of these harmonics to

define our function,

$$f(\theta,\phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta,\phi). \tag{2.70}$$

In this case, $a_{\ell m}$ are the coefficients and are analogous to the Cartesian function in Fourier space and thus are defined in a similar manner

$$a_{\ell m} = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} f(\theta,\phi) Y^*_{\ell m}(\theta,\phi) d\Omega, \tag{2.71}$$

where Ω is the solid angle and $Y^*_{\ell m}$ the complex conjugate of the spherical harmonic. As with the Fourier transformed functions above, we can define a power spectrum $C_(\ell)$ as the variance of the coefficients.

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell. \tag{2.72}$$

The power spectrum in this form can be written as $C_\ell = (\sum_{m=-\ell}^{\ell} \langle |a_{lm}|^2 \rangle)/(2\ell+1)$ and so is the average squared magnitude of the coefficients for each ℓ. But as one can see there is only $2\ell+1$ m-modes for each multipole leading to the following cosmic variance $\delta C_\ell = \sqrt{2/(2\ell+1)}$.

Returning to the shear field, which can now be written in terms of spherical harmonics,

$$\gamma(\theta,\phi) = \sum_{\ell m} (\varepsilon_{\ell m} + i\beta_{\ell m}) Y_{\ell m}(\theta,\phi), \tag{2.73}$$

where $\varepsilon_{\ell m}$ and $\beta_{\ell m}$ are the multipole coefficients of the E-modes and B-modes, respectively. We can then relate the E-mode angular power spectrum to the lensing potential angular power spectrum,

$$C_\ell^{\varepsilon\varepsilon} = \frac{1}{4} \frac{(\ell+2)!}{(\ell-2)!} C_\ell^{\phi\phi}, \tag{2.74}$$

where $C_\ell^{\varepsilon\varepsilon}, C_\ell^{\phi\phi}$ are the E-mode angular power spectrum and lensing angular potential spectrum respectively. this relation is possible due to the relationship between the E-modes and the lensing potential in the first place being $\varepsilon_{\ell m} = \sqrt{(\ell+2)!/(\ell-2)!}\phi/2$.Much in the fashion seen previously, we can now link the lensing potential angular power spectrum to the matter power spectrum just as we had done with the convergence before,

$$C_\ell^{\phi\phi} = \frac{8}{\pi} \left(\frac{3\Omega_m H_0^2}{2c^2}\right)^2 \int \frac{dk}{k^2} I_\ell^2(k) \int d\chi \left(\frac{\chi' - \chi}{\chi' \chi a}\right)^2 j_\ell^2(k\chi) P_\delta(k,\chi), \tag{2.75}$$

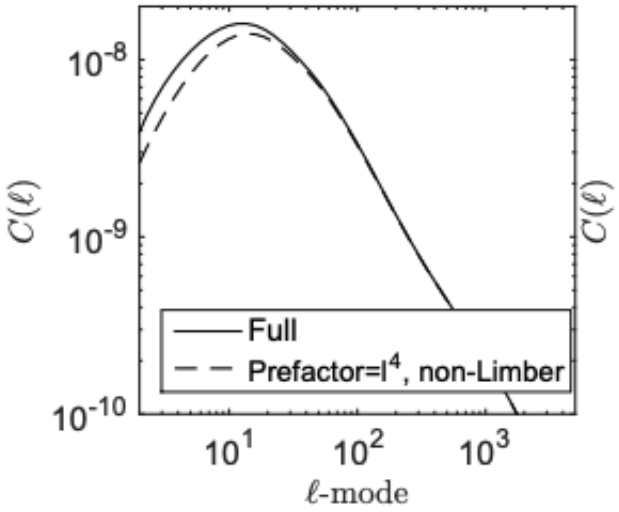


Figure 2.7: The difference in observed matter power spectrum between the full version (bold line) and that that has used the 'flat sky' approximation (Kitching et al., 2017).

where

$$I_\ell(k) = \int d\chi \left(\frac{\chi' - \chi}{\chi' \chi a}\right) j_\ell(k\chi) [P_\delta(k,\chi)]^{1/2}, \tag{2.76}$$

where j_ℓ is the spherical Bessel function. For full consistency, the shear correlation functions can also be expressed in terms of the shear E-mode angular power spectrum as

$$\xi_\pm(\theta) = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell^{\varepsilon\varepsilon} d^\ell_{2,\pm 2}(\theta), \tag{2.77}$$

where $d^\ell_{2,\pm 2}$ are the reduced Wigner D-matrices.

Both equation 2.75 and equation 2.77 are examples of the exact equation without any of the approximations (minus the Born approximation, which was mentioned above and allowed for the E-mode, lensing potential relation). However, this full term is tough to calculate in reality, so the flat sky approximation is introduced. The flat sky approximation allows us to neglect the angular nature of our observations and instead presume the observations took place on a euclidean surface. This assumption may be allowed, especially when the observation field has taken place over a small angular change (and thus is also known as a small-angle approximation). Mathematically this replaces our expansions in spherical harmonics with the equivalent expansions in Fourier modes. Due to the similar orthogonality relations between spherical harmonic functions and exponential functions (as the sine waves can be represented as exponential functions), the physical changes to equation 2.77 are minimal. Firstly the preface in the relationship between E-modes and lensing potentials in

equation 2.74 simplifies to $(\ell+2)!/(\ell-2)! \to \ell^4$. This is due to the different ways spin raising and lowering operations act on harmonic and exponential functions. The underlying effect is the replacement of the reduced Wigner D-matrices with the Bessel functions used prior i.e. $d^{\ell}_{2,\pm 2}(\theta) \approx J_{0/4}(\theta\ell)$. To test the appropriateness of this approximation (Kitching et al., 2017) ran simulations of the cosmic shear power spectrum for a given CFHTLenS matter distribution using the cosmology equal to the 2016 Plank results, obtaining both the full cosmic shear power spectrum and that power spectrum one would get from a flat-sky approximation. From Figure 2.7 we can see the power spectrum and their relative differences. It is clear that on large scales $\ell < 10$, there is a more than 10% suppression in power due to the Flat-sky approximation, a suppression that reduces to $< 1\%$ for small scales $\ell > 100$. This is expected as the approximation assumes a small angle of observation, thus a small scale. The same is true when considering changes to the correlation function, with effects only tangible at high angles. It is also worth noting that the approximation has a much larger effect on ξ_+ than on ξ_- which is to be expected as ξ_- is more localised (thus waiting for smaller scales more than larger scales) than $\xi+=$ due to the J_4 Bessel function as discussed earlier. When used on its own, the approximation has met criticism due to its inducing errors of the order $1/\ell$, which may not be completely negligible.

Another issue that makes calculation difficult and time-consuming to evaluate accurately at high multipoles is the rapid oscillations in the spherical Bessel function. To attack this function, we use the Limber approximation (Limber et al., 1953), and its generalisations in Fourier space (Kaiser et al., 1992). The approximation assumes that the rest of the integrand in $I_\ell(k)$ is slowly varying compared to the variations in j_ℓ which is generally the case for large ℓ. This assumption allows us to replace the spherical Bessel function with a much more practical delta function i.e. $j_\ell \to \sqrt{\pi\delta_D(\nu - k\xi)/2\nu}$. Where $\nu = \ell + 1/2$. The wave number k is then related to the radial distance ξ via the relation $k\xi = \nu$. The practicality of this approach is that it allows us to effectively ignore the second integral so that when combined with the flat sky approximation, we can return to the original equation. A more digestible explanation of the approximation is given in (LoVerde & Afshordi 2008), which says that it assumes the matter power spectrum is not evolving. As radical of an assumption as this may be, it can be shown that applying the Limber approximation to a spherical-radial representation of the cosmic shear field can lead to a first-order approximation of the full cosmic shear power spectrum despite the constant matter power spectrum assumption.

To assess the full impact of the approximations (Kitching et al., 2017) computed the integrated effects over the differences using the following formula:

$$\langle A \rangle / N_A = \frac{\int d\ln(\ell)\ell^2 \delta C(\ell)}{\int d\ln(\ell)\ell^2}. \tag{2.78}$$

Similar formulas are provided in (Massey et al., 2011), (Laureijs et al., 2011) but also included a normalisation term $N_A = \int dln(\ell)\ell^2$. The use of the formula is that non-zero values will change the power spectrum amplitude and lead to biases in cosmological parameter inference. Through the simulation, they found that the flat-sky approximation combined with the limber approximation led to an amplitude difference of $\langle A \rangle / N_A = 1.9 \times 10^{-13}$. For context, the requirements for a Euclid or LSST-like weak lensing survey are for that quantity to be $\langle A \rangle / N_A \leq 1.9 \times 10^{-12}$ across all systematic effects (instrumental and approximations in the mathematics) to produce unbiased results for dark energy equation parameters. This means that the Flat sky and limber approximation would account for 11% of the entire systematic effects budget and thus is way too high to be useful in the future. With that said, a common counterargument is that the approximations were only meant for the small scales of which small angel approximations are used; however, if we limit our selfs to said small angles and thus ignore scales larger than $\ell < 100$, then the formula produces a bias of$\langle A \rangle / N_A \leq 1.7 \times 10^{-12}$ which is still a high percentage of the total budget. These approximations are still used today due to their practical and easily calculable nature but are slowly being phased out due to their lack of necessity as computational power increases, and error budgets continue to fall.

2.5 Cosmological Analysis

2.5.1 Estimation

At this point, we have methods for estimating the ellipticity of galaxy images collected in comprehensive surveys (be it weighted Quadrupoles moments, shapelets, or other methodology). We also have equations that allow us to estimate the shear correlation function based on this ellipticity. Furthermore, these same correlation functions can be converted to shear power spectra and thus convergence power spectra. Lastly, through the above approximations (mainly Flat-sky and limber), we can relate these convergence power spectra estimations to the universe's matter distribution. With this framework, we can now link observation to the theoretical prediction, finally allowing us to finally make some actual cosmological predictions.

Weak lensing is so cosmologically valuable because it is a direct probe to the whole matter distribution, as opposed to the other methods that rely on baryonic tracers such as Lyman-α and galaxy clustering. These tracer methods have to deal with considerable complications due to the non-linear aspects of galaxy bias and redshift-space distortion. That is not to say weak lensing is without its issues. First, even if we neglect all the uncertainty that

comes with the approximations and estimations needed to go from image to modelled matter distribution, there is the fact that weak lensing by nature is 'weak', and so millions of galaxy images are required for significant enough shear signals. This is a technological limitation as weak lensing surveys continue to increase in their observation volume. The second issue is that weak lensing provides a projected matter density field, a combination of all the matter between source and observer. The shear field is 2 dimensional. This flatness means weak lensing is insensitive to many available modes of the complete matter distribution and thus leads to a high information loss.

Some methods are attempting to mitigate this information loss. One such method is known as 3D lensing (Heavens, 2003). In this method, one does not only take elliptical information from galaxy images but redshift information as well (usually in the form of photometric redshifts, but this will be discussed in more detail below). The individual redshift information is used so that when a spherical harmonic transformation of the shear is made, one does not need to mix the k-modes along a line of sight (which happens in an average projection) but instead has a shear field consisting of both ℓ-modes and k-modes for the point in the sky and the line of sight respectively. Through simulation, it has been shown that by taking a 3D approach, one may be able to estimate the equation of state cosmological parameter ω to up to 1% precision. one may even be able to constrain the equation of state across different redshifts ($\omega(z)$) which allow for the testing of different vacuum energy models. Furthermore, such analysis could reduce error bars in matter distribution parameters such as amplitude with even modestly accurate photometric redshifts. Other advantages include eliminating several systematic effects such as intrinsic alignment. However, a concern for this method is the contamination of cosmological analysis due to survey masks (areas of no observation) due to the complete spherical harmonic transformation of the field masked areas' course power to mix between mask-affected and unaffected scales. With this in mind, the geometry of the mask has to be accounted for, which can be achieved in several ways.

1. applying the effect of the mask directly onto the theoretical prediction through a mixing matrix. One such method for being pseudo-$C_\ell s$ (Hivon et al., 2002)

2. by correcting the data for the mask through a regularisation scheme

The second suggestion requires a very intensive matrix inversion. Either way, surveys need to be incredibly large to account for masks and thus have only been a viable form of analysis recently, with the CFHTLens (Kitching et al., 2014) survey being 120 square degrees. However, as with the 'weak' issue, this issue will decrease as surveys get larger over time.

An alternative, simpler (though potentially less effective) method that is the current method of use across most modern surveys is tomographic redshift binning (Hu, 1999). In this approach, redshift data is used again, but instead of caring about every individual galaxy's photometric redshift for the full shear analyses, the survey is split into N_z tomographic bins of which a galaxy is allocated to a bin based on its photometric redshift. This means that one no longer obtains one convergence power spectrum but $N_z(N_z-1)/2$ convergence power spectra $P_{\kappa,ij}$ where $1 \leq i \leq j \leq N_z$ as both auto spectra $i = j$ and cross spectra $i \neq j$ are included. This change in analysis modifies our convergence power spectra equation somewhat:

$$P_{\kappa,ij} = \frac{9}{2}\Omega_m^2(\frac{H_0}{c})^4 \int_0^{\chi_{\rm lim}} d\chi \frac{q_i(\chi)q_j(\chi)}{a(\chi)} P_\delta(k = \frac{\ell}{f_k(\chi), \chi}), \tag{2.79}$$

where $q_i(\chi)$ is the lensing efficiency for the i'th bin defined as

$$q_i(\chi) = \int_\chi^{\chi_{\rm lim}} d\chi' n(\chi')_i \frac{f_K(\chi' - \chi)}{f_K(\chi)}. \tag{2.80}$$

In this case, $n(\chi)_i$ is the source galaxy distribution of the i'th tomographic bin such that $n(\chi)_i d\chi = n(z)_i dz$ the lens efficiency indicates how strong the lensing effect it as distance χ due to the combined background distribution. There is a difference in strength between the galaxy distribution and the lensing efficiency as a function of distance, the lensing efficiency is a very broad function in comoving distance or redshift; thus, tomographic power spectra are not independent of each other. This means that a large-scale structure, regardless of the bin it is in, can contribute to all the bins with higher redshifts as the light from the source to the observer would also have to travel between those bins and interact with the large-scale structure. Even if this were not the case, photometric redshift errors mean galaxies can be placed in the wrong tomographic bin, leading to overlapping tomographic redshift distributions.

2.5.2 Prediction

Through such observations as the change in the shape of observed galaxies, one can use techniques such as tomographic bins to build up an estimated matter power spectrum. However, as seen previously, one can not infer information from the universe from this observation alone; instead, it must be compared to models of how we believe a matter distribution and thus a lensing distribution would evolve with redshift, and through these comparisons, we can predict the value of specific cosmological parameters. With that said, weak lensing is incredibly limited in the cosmological parameters it can probe, even among

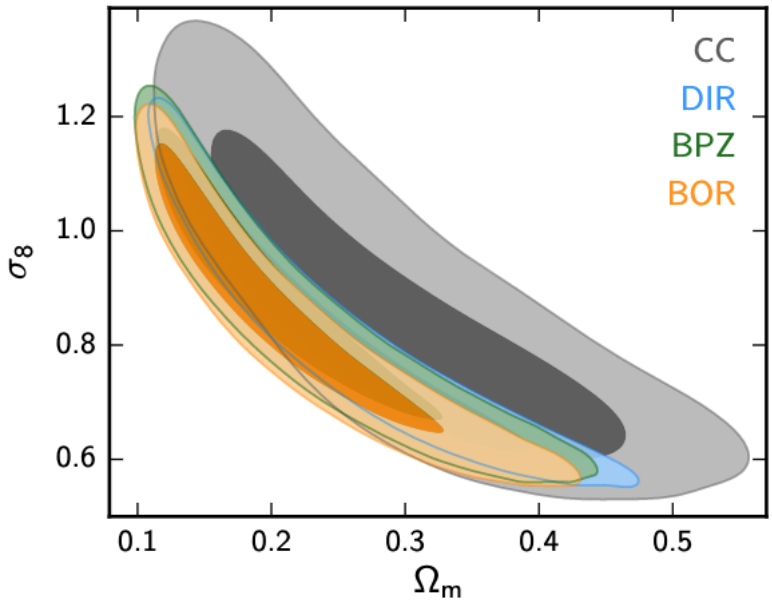


Figure 2.8: The cosmological constraints on values of Ω_m and σ_8 using different methods for inferring redshift distributions of weak lensing surveys (Hildebrandt et al., 2017).

those that are intuitively connected, such as Ω_m and σ_8. Take, for example, the results from the Kilo Degree Survey that observed 449 deg^2 of the sky (KiDS-450 (Hildebrandt et al., 2017)) as seen in Figure 2.8. Using images of 14640774 objects assorted into 5 tomographic bins, shear correlation functions were compared to models. From Figure 2.8 we can see multiple banana shapes, all representing possible samples of the actual cosmological values using different redshift distribution techniques (see Chapter 3). No matter the technique, one still gets an effect, meaning parameters scale across multiple values. This is due to a degeneracy they both hold. Different combinations of both parameters can cause identical shear power spectra to be observed. This makes intuitive sense. One would expect that a very clumpy but not very dense universe would deliver similar levels of weak lensing to that of a dense but more spread-out universe.

Due to this degeneracy in result, it has become a lot more common to adopt a cosmological parameter that combines the 2, known as S8 where $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$. Visually this combination of the 2 variables represents the central 'width' of the banana and allows for a visual uncertainty contour that is much more adept for a fundamental Gaussian uncertainty, as seen in Figure 2.9. These are updated results not only including KiDs-450 as seen above but other extended survey results from the KiDs-VIKING survey, as well as comparisons to other weak lensing surveys such as the Dark Energy Survey (DESy1 purple), the Hyper Suprime-Cam Subaru Strategic Program (HSC orange). These particular results are all

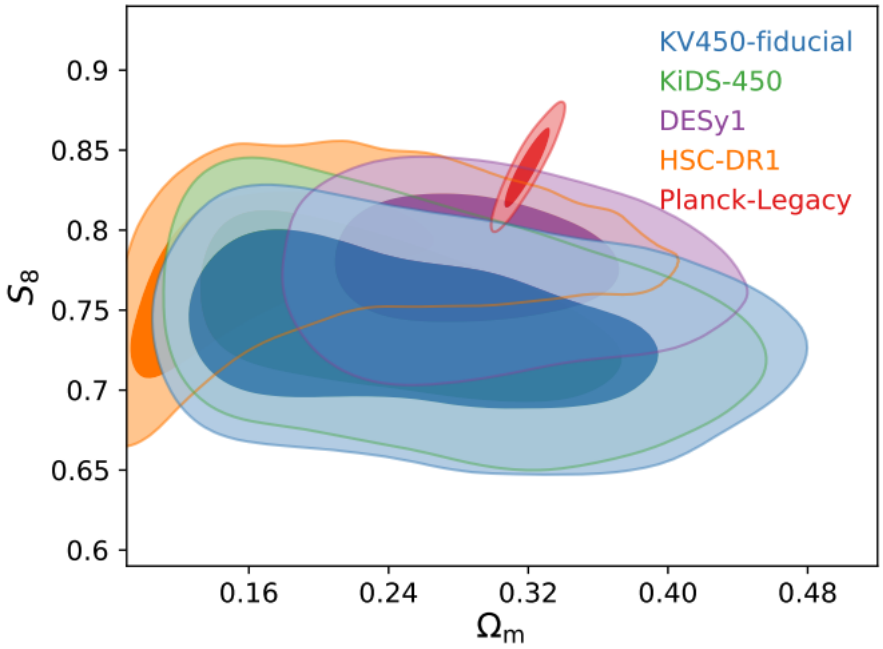


Figure 2.9: Cosmological constraints on parameters Ω_m and S_8 for different weak lensing (and CMB) surveys (Hildebrandt et al., 2018)

broadly in agreement, but one result reveals one of the other central tensions of cosmological physics.

2.5.3 Challenges

The predictions mentioned above make weak lensing a powerful tool for cosmology, especially in parameter constraints. However, the above theoretical explanations for the link between cosmological matter distributions and observable shear correlations missed some complex challenges that still face weak lensing going forward.

One of the major issues is that we have presumed that all shear correlation observed through our weak lensing surveys is caused directly and exclusively by the underlying matter distributions between source and observation. In other words, we have assumed that the source ellipticity of the galaxies is entirely randomly generated so that the average shear $\langle\gamma\rangle$ is zero on large enough scales. Unfortunately, this is not the case. The ellipticity of galaxies is not completely randomly generated but instead has some correlation to other galaxies' shear before weak lensing. They have intrinsic alignment. This alignment works to contaminate the shear signal. It can lead to the signal being boosted by the intrinsic alignment

between galaxies that mimic the gravitational shear or the complete opposite. Galaxies can have intrinsic anti-correlation diminishing the shear signal (Kiessling et al., 2015).

The believed reason for this intrinsic alignment is due to gravitational fields from structures more immense than the galaxies coursing tidal interactions. As galaxies can vary in size, so can structure coursing the alignment; thus, intrinsic alignment happens simultaneously on multiple scales, from large halo-sized scales above 10Mpc to typical galaxy-size scales below 1Mpc. The alignment is usually split into 3 scale regimes, the 2 mentioned above and an intermediate galaxy cluster scale of a few Mpcs. These regimes will be covered in more detail below.

On the largest scale, we consider the intrinsic alignment of the galaxy is caused by the dark matter halo in which they are. To the lowest order particles at position $x(q,a)$ where q and a are initial position and scale factor receptively. These particle positions follow straight lines determined by the gradient of the displacement potential Ψ (Zel'Dovich et al., 1970)

$$x(q,a) = q - D_+(a)\delta\Psi(q), \tag{2.81}$$

where D_+ is the growth factor. We can now consider the entire halo by taylor expanding the above equation along the halos centre of gravity $\bar{q}$,

$$x(q,a) \approx q - D_+(a)[\delta\Psi(\bar{q}) + \sum_j (q_j - \bar{q}_j)\delta\Psi(q)]. \tag{2.82}$$

The two terms within the square brackets identify the difference in halo motion. In the first term, I identify the bulk motion of the halo, while the second term shows the difference in motion of a particle compared to the centre of gravity. The field affecting this motion difference is the title field tensor $\delta\Psi(q)$. For example, suppose the tidal field of the halo does not coincide with the halo's principal axis. In that case, a shearing motion produces angular momentum, of which an individual galaxy galactic disk would form perpendicular to the halo angular momentum direction coursing a quadratic alignment for spiral galaxies. This effect is also known as tidal torquing, and by tracking the ellipticities of spiral galaxies, one can reasonably predict the halos angular momentum direction but not necessarily its magnitude (Catalan and Theuns, 1996). In terms of its effect on contaminating weak lensing signal, (Capranico et al., 2013) computed the angular ellipticity power spectra and found that the intrinsic E-mode spectra caused by quadratic alignment were an order of magnitude less than the weak lensing power spectra, so the minimum contamination but still valuable for identifying systematics as it did produce a B-mode signal as well.

In the case of elliptical galaxies, the galaxy structure is determined by the velocity dispersion of dark matter particles and starts within it so it can reach viral equilibrium. However, if

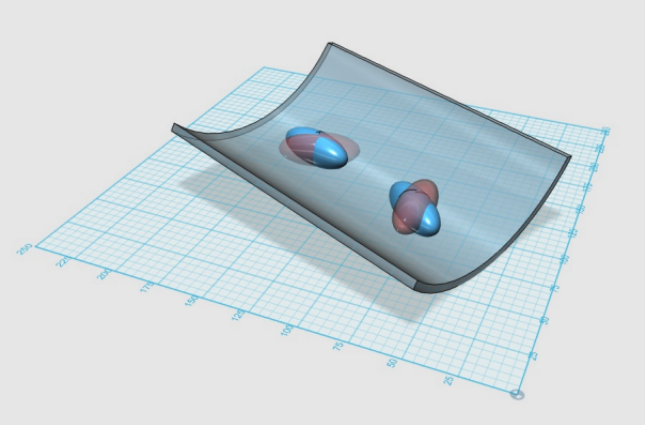


Figure 2.10: 2 elliptical galaxies within a halo's tidal field (grey sheet) shape will change from uncorrelated (red) to correlated (blue) (Kiessling et al., 2015).

said galaxy is placed within a tidal gravitational field, it will distort, forcing the stars to take new positions to keep the equilibrium. These new positions will lead to a new galaxy shape that correlates with other nearby galaxies within the same tidal field, as shown in Figure 2.10. In the limits of weak fields, the relationship between elasticity and the projected tidal shear field is entirely linear, allowing for the following model for intrinsic shear (γ^I):

$$\gamma^I = -\frac{C_1}{4\pi G}(\delta_x^2 - \delta_y^2 + 2i\delta_x^2 y)S(\Phi), \tag{2.83}$$

where x, y are Cartesian coordinates in the sky plane and $S(\Phi)$ is a smoothing of the gravitational potential on the halo scale. C_1 is a constant of proportionality that indicates the strength of the reaction the galaxy feels as its shape changes due to the tidal field. Through measuring ellipticity correlation in the SuperCOSMOS (Brown et al., 2000) dad the constant was found to be $C_1 \approx 5 \times 10^{-14}\,(h^{-1}\,M_\odot\,\mathrm{Mpc}^{-3})^{-1}$ (Bridle and King, 2007). The linear relationship is not limited to particular constrained conditions either. The range in which the linear relationship holds is quite high linear ellipticity correlations have shown to extend to $30h^{-1}$ Mpc (Okumura et al., 2008) in the Sloan Digital Sky Survey and extended as far as $100\,h^{-1}$ Mpc for cross-correlations between galaxy shape and density (Mandelbaum et al., 2006). With this said, studies have found that the linear model alone is strong enough to account for the total intrinsic alignment observed in elliptical galaxies (Camelio and Lombardi, 2015).

As intrinsic alignments on the post-halo scale cannot account for all observed intrinsic alignments, the sub-halo scale must also be considered. Under these circumstances, one can make intrinsic alignment predictions through the halo model itself. Most important, through the halo model, one can differentiate the alignment of central galaxies from the

alignment of satellite galaxies. It is often assumed to simplify the model by stating that central galaxy ellipticity was uncorrelated with satellite ellipticity. Matter can be further simplified by using the fact that the dominant term for galaxy alignment is the same term that describes the alignment of their host halo. This means one can account for the total halo alignment power spectra using only a few terms. In this state, the 2 models shown above (the small-scale Halo model and the large-scale tidal interaction model) can not be extended to their intermediate regime for the small scale because the model relies on the halo is self so bound to that scale while the large scale one is derived from Lagrangian perturbation theory which can only be applied to potentials from large scale structure. Figure 2.11 shows the issue more acutely, included in the observations of galaxy correlations found in SDSS luminous red galaxies at various scales. Also included are different models. Only the one Halo model correctly models the observations at the most miniature scale, while only the linear models can fully account for the most significant scale. From this, one may look for an intermediate scale theory that joins the 2 scales to model the intrinsic alignments of galaxies fully. These scales are sometimes referred to as mildly non-linear as observations have shown that intrinsic alignment seems slightly more potent than the linear predictions. Methods have been proposed, such as phenomenological solutions or replacing the matter power spectrum with a non-linear counterpart, this model named Non-Linear Alignment (NLA) gives a more substantial predicted alignment at more minor scales while keeping things unchanged at larger scales (one can see from Figure 2.11 that the NLA model -red-bridges the gap between large and small scales but still fails at the smallest). However, as shown through their fails and lack of overlap in specific scales, these are toy models, and thus a complete understanding of intrinsic alignment has not yet been met. With a complete understanding, not knowing intrinsic alignment must be considered within the weak lensing analysis itself.

To analyse weak lensing, it is essential to gauge its effect on core mathematics. We can expand upon what we mean by observed ellipticity so that it follows $\varepsilon = \gamma^G + \gamma^I + \varepsilon^s$ where ε^s is the source ellipticity. It is unlensed, random and presumed to be sampled from a distribution that averages to zero. γ^G is the shear we have been familiar with up to this point; the shear is caused by a gravitational potential, standard weak lensing. The new extra term is γ^I which contributes to the average ellipticity based on its environment. Notice that intrinsic alignment is not treated under the source ellipticity variable under this equation but is instead treated as its kind of shear being 'applied' to the source. Under this consideration, one can work out the ellipticity power spectrum to find 3 terms where there used to be 1:

$$C_\ell^{\varepsilon\varepsilon} = C_\ell^{GG} + C_\ell^{GI} + C_\ell^{II}. \tag{2.84}$$

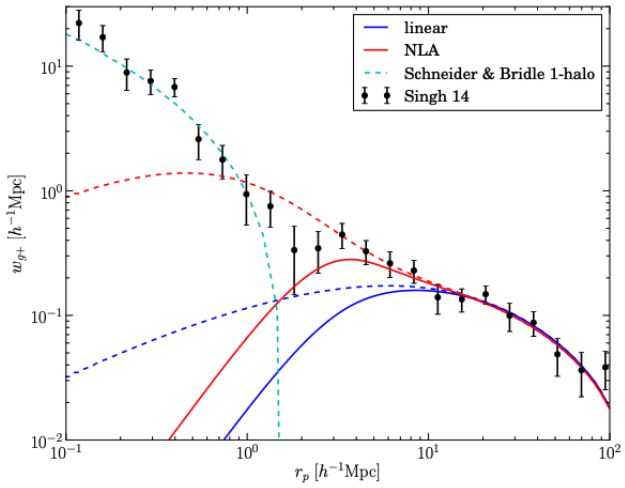


Figure 2.11: 2 elliptical galaxies within a halos tidal field (grey sheet) shape will change from uncorrelated (red) to correlated (blue) (Kiessling et al., 2015).

The three components contribute to the weak lensing signal in different ways.

- The first term C_ℓ^{GG} is the usual weak lensing power spectrum one has seen up to this point where the ellipticity of the source is correlated due to mass structures between source and observer
- The last term C_ℓ^{II} represents the correlation one would expect to be caused by intrinsic alignment where 2 nearby galaxies have similar ellipticities due to being mutually aligned by the gravitational tidal field
- The second term C_ℓ^{GI} represents a correlation caused by both the intrinsic alignments and the weak lensing effect. For example, consider a sizeable cosmological structure with a large gravitational tidal field. Any galaxies near it will be intrinsically radially aligned to it, whereas any galaxies behind the large structure will be tangentially aligned to the structure due to weak lensing. This effect leads to the 2 galaxies in question being anti-correlated.

The II and GI effects are well presented in the illustration of Figure 2.12 where in both cases, we have a large mass structure and galaxies being intrinsically aligned to it at redshift z_i, but we also have a galaxy being gravitationally lensed by the structure at redshift z_j where $z_j > z_i$. One should note, though, that it is only in the simplest of alignment models do intrinsic alignment and gravitational shear act orthogonally coursing an anti-correlation. More complex models that account for the merger history of galaxies or baryonic infall may result in different intrinsic alignments coursing positive correlations with lensing. It is also

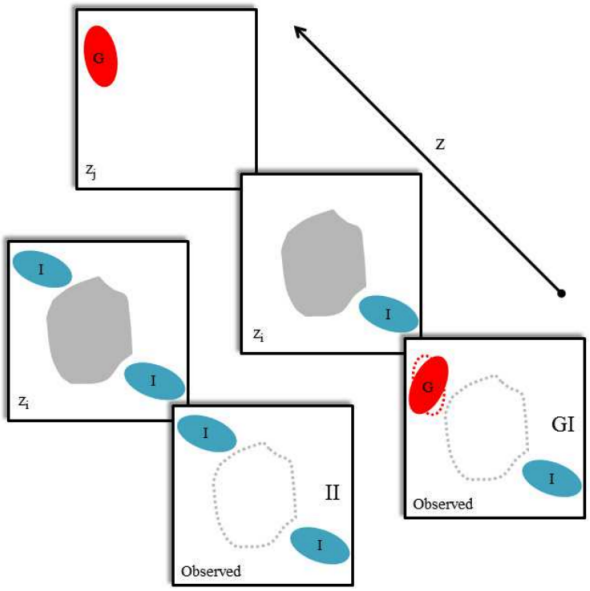


Figure 2.12: (Troxel and Ishak, 2014)

worth noting there is one omission to the terms impacting the elliptical power spectrum, and that is C_ℓ^{IG} this refers to the correlation between galaxies that have been intrinsically aligned by a large mass structure and the galaxies in front of that structure that have been weak-lensed by it. However, this term can be rightfully ignored as galaxies are only lensed by large mass structures between source and observer, not behind the source.

To mitigate the effect of intrinsic alignment, one can take a direct approach of including a parameterised model of intrinsic alignment and marginalising its parameters as a part of the cosmological analysis. The parameterisation may be more general and be a function of galaxy properties such as redshift and galaxy type that can also be marginalised as new parameters. One such early example was developed as King and Schneider (2003) who had the intrinsic correlation function as a linear combination of 9 intrinsic alignment functions parametrised as $\eta(r,z) = (1+\bar{z})^\alpha \exp(-r/R)$ where r is the comoving separation, R is the correlation length and $\bar{z}$ is the mean redshift of the galaxy pair. In this case, 3 sets of (α, R) are the nuisance parameters to be marginalised. Using this approach, the degeneracy between Ω_m and σ_8 and future works showed a similar effect for including the GI contribution (KING2005). There are, however, major limitations to this approach. As we saw earlier in Figure 2.11, there currently is no model that succeeds under all scales. Thus, one standard model cannot be assumed for the whole cosmological analysis; thus, if one cannot trust the model on certain scales, one would be unable to trust the cosmology inferred in and out of those scales.

An alternative approach is to remove the intrinsic alignment data from the analysis in a process known as nulling (Joachimi and Schneider, 2009). Intrinsic alignment is directly linked to redshift. The closer the 2 galaxies (thus, the closer their redshifts) are to each other, the more likely they are to exist within the same gravitational tidal field and have the exact intrinsic alignment. One can exclude or down weight 'close' galaxies, thus significantly reducing the magnitude of the II power spectrum. However, this has the apparent cost of reducing the overall number of galaxies pairs and thus the statistical power of the weak lensing signal, which can outweigh the benefits. (Takada and White, 2003) showed that applying the nulling technique to a topographic survey of 5 or more redshift bins (so only using the cross-correlations between the tomographic bins) would increase errors on parameter constraints by about 10%. In contrast, the error caused by II contamination would be negligible. In terms of GI, the removal of certain data is also possible. Limiting the galaxy data by type one can exclude galaxies that contribute primarily to the signal. Such a type difference exists. CFHTLenS confirmed that the intrinsic alignment signal from late-type galaxies is consistent with zero while early-type galaxies (which have had longer to align) course a non-zero GI signal. It has been shown that removing 20% of red foreground galaxies can suppress the intrinsic alignment signal by 2. There are then 2 main floors to this approach. The first-mentioned above is the reduction in statistical power; even with perfect redshift information and where one can confidently use 20 or more redshift tomographic bins, even though the GI signal can be removed entirely, the size of the 2d confidence regions in parameter estimation will increase by 20% - 50% (Joachimi and Schneider, 2009). This is an ideal scenario. In reality, insufficient photometric data is used, leading to galaxies being missed placed in bins leaking the intrinsic signal. Under these conditions, the increase in contour size is still around 50% while the capability to reduce the intrinsic alignment signal is cut severely (Joachimi and Schneider, 2009).

With this in mind, there is a third approach that may be applied to try to reduce intrinsic alignment signals. This approach is known as the self-calibration technique. The approach gets its name through using observables within the same survey and re-scaling their cross-correlations to calibrate the weak lensing survey. In this case, the extra observable is the galaxy density power spectra (denoted as C^g) which allows us to examine the shear-density cross-correlations and density-density correlations (Fry and Gaztanaga, 1993). The technique also uses the $i < j$ redshift separation examined in the nulling effect. This separates from nulling because of the inclusion of the density observable. This approach differs slightly if we self-calibrate using the cross-correlation between tomographic bins within a single tomographic bin.

The II signal is negligible between tomographic bins where the $i < j$ nulling rule is applied. This leaves us with the IG signal, which can be related to the density-intrinsic alignment cross-correlation

$$C_{ij}^{IG}(\ell) \approx \frac{W_{ig}^{G}}{b_1^i \Pi_{ii}} C_{ii}^{Ig}(\ell), \tag{2.85}$$

where b_1^i is the average galaxy bias within the ith redshift bin under the assumption that galaxy and matter density can be linked through a galaxy bias function such that

$$\delta_g(x) = b_1(x)\delta_m(x;\chi) + \frac{b_2(x)}{2}\delta_m(x;\chi)^2. \tag{2.86}$$

Although we are neglecting the b_2 terms, the extra bias this non-linear term would have caused leads to inaccuracies of less than 10% for a typical Stage IV weak lensing survey. In weak lensing signal terms, this reduces GI contamination by 10 or more (Zhang, 2008). This is similar to the reductions we saw in nulling techniques; however, the inclusion of self-calibration holds its only advantages. For one, the GI signal is not simply eliminated in nulling but simultaneously estimated. The benefits of this are 2 fold, one can you the estimated GI signal for other alignment-based studies, but also one can use the signal to preserve statistical power and not have the increase in contour size seen above due to statistical reduction. There are issues with the lack of elimination of GI, and an estimator must be used to separate the lensing and intrinsic alignment components o the shear-density cross-correlation. As this separation is imperfect, an extra error is propagated into the parameter estimations. The method also relies on the assumption of galaxy bias; thus, a bias model must be measured as well; however, the errors in uncertainty caused by galaxy bias inaccuracy would be negligible for the biases ensured in stage IV surveys (Zhang, 2008).

A similar technique is used in the case of self-calibration within a photometric tomographic bin. However, in this case, instead of using a single scaling relationship as seen in equation 2.85, a set of scaling relationships are constructed between the shear and intrinsic alignment and the shear and galaxy density cross-correlations. All these relations have almost identical redshift separation dependencies, so various measurements of shear and density cross-correlations are taken with micro bins of width 0.01z across single bins. The micro bins may alert the issue of photometric error, but the method was shown to be safe under the redshift errors and catastrophic errors expected of stage IV surveys. The approach can also be modified only to give individual signals if that is what has required allowing for more flexibility. In both self-calibration methods, the technique can be expanded to 3-point statistics allowing for insights into the bispectrum and non-gaussian distributions.

A promising method allows us to not self-calibrate at all. Instead, one may use to calibrate using external datasets. One example is the cross-correlation between CMB gravitational lensing and galaxy lensing (Vallinotto, 2011). When first detecting the CMB lensing - galaxy lensing cross-correlations was an exciting discovery. Even though the CMB is unaffected by intrinsic galaxy alignments, a GI signal was still detected by Hand et al. (2015) (although to separate from usual galaxy-based GI, it is labelled ϕI (Troxel and Ishak, 2014)). This is due to the intrinsic ellipticity of foreground galaxies correlating with the lensing deflections caused by the CMB temperature fluctuations and polarisation signals. The observed signal can then be described in terms of its components

$$C_i^{obs} = C_i^{\phi G} C_i^{\phi I}. \tag{2.87}$$

A combination of linear alignments models that agree with intrinsic alignment measurements at low redshift studies has shown that the ϕI signal is about 15% the size of the ϕG signal making its contaminating impact 50% more than that of GI contamination on galaxy lensing (Troxel and Ishak, 2014). At first, this seems like a step back, but it has also been shown that linear scaling relationships can be made between IΦ, IG and Ig. Furthermore, these relationships are more accurate than that seen in equation 2.85 with a typical error of 5% for stage IV surveys. Although requiring an overlap between the galaxy and CMB survey measurements, this method gives an additional study of the ig signal as in a proper focus of weak lensing surveys going forward.

One may wonder why the separation of E and B modes cannot be taken advantage of. As stated earlier, shear signals are related to the gradient of the gravitational potential. Thus, they should be curl-free (consisting only of E-modes), so any detection of a B-mode could indicate intrinsic alignment (as I signals do have E and B modes). This indication has led to suggestions that the B-mode signal can lead to constraints on the contamination (Heymans and Heavens, 2003) or as diagnostic tools to test the above methods for eliminating the I signal altogether. However, in reality, the usefulness is limited due to the lack of correlation between the I B-mode component and the G E-mode component in the GI correlation. Furthermore, the leading terms from the decomposition predict no B-mode signal further limiting the diagnostic use.

Lastly, there have been suggestions that through other novel detections, we can have an idea of what the intrinsic shape of a galaxy was independent of the weak lensed image we gain from extensive surveys. Such a detection would allow us to decouple intrinsic alignment and shape noise from the weak lensing signal, increasing our signal-to-noise ratio. Such an example of decoupling of this nature would be to look at the radio signals of a galaxy image to determine the polarisation emission of the galaxy (Brown and Battye, 2010). The

advantages of determining the polarisation are that they are unaffected by the lensing effect and give insight into the intrinsic alignment of the galaxy. A similar approach can be taken with optical polarisation (Audit and Simmons, 1998). With that said optical signals are more easily affected by external factors such as other source galaxies and intervening magnetic fields, which introduce position dependencies requiring extra knowledge to overcome. An alternative approach requires an overlapping smaller spectroscopic survey in the region of the extensive photometric weak lensing survey (Huff et al., 2013). This would allow one to measure the rotation velocity of disk galaxies. Such information could be combined with the Tully-Fisher relation (Tully and Fisher, 1977) which describes a relationship between rotational velocity and luminosity of disk galaxies to identify the inclination of the said galaxy. Through this inclination, one can find the intrinsic alignment of the galaxy before gravitational lensing. In both cases, the statistical power of the weak lensing surveys could be increased by these decoupling methods, but they require extra resources either in whole new complimentary surveys or extra detection in specific frequencies. With that said, intrinsic alignment stays as one of the main factors in limiting weak lensing signals, a factor that is being approached and limited in many different ways, as shown above.

2.6 Summary

To summarise, this chapter looked into the final probe for insight into the cosmological universe. The phenomenon of gravitational lensing is the idea that photons will change directory and bend due to the presence of mass. A phenomenon that predates Einstein and can be predicted (though inaccurately) from Newtonian physics. Through Fermat's principle and the borne approximation, one can create a Jacobian matrix that takes photons' changing trajectories and maps the distortion these changes course in an image. These distorted images lead to variables that make up the Jacobian, the convergence and shear, two variables that can be interpreted from the shape of the galaxy image, most notably the ellipticity of the galaxy image. In fact, in the regime of weak lensing where $|\gamma|, \kappa < 1$, the average ellipticity of galaxies becomes an unbiased estimator of the reduced shear $< \varepsilon >= g$ under the assumption that a statistical galaxy orientation is entirely random so that without weak lensing having an effect the mean ellipticity would generally be 0. Ellipticity cannot be easily observed in most cases due to PSF, so methods were considered to estimate the ellipticity under these conditions. Under the assumption that ellipsis and thus shears can be accurately detected and through methods such as tomographic binning power spectra can be constructed and linked to the matter power spectrum. This links cosmological parameters such as Ω_m and mass density of the universe, and σ_8 the clumpiness (although both are usually combined into one S_8

parameter) of the universe can be predicted and compared to other independent observations. Most notably, the tension between CMB and weak lensing σ_8 estimation. Challenges do still exist in this study of work, most notably with the underlying assumption that galaxies' orientations are statistically random. This is not the case as the universe's growth can lead to galaxies having intrinsic alignment contaminating the weak lensing signal, but many methods exist to compensate, reweight or completely disentangle the intrinsic alignment effect. With that said, one of the main contributions to an error within weak lensing estimation has yet to be mentioned in full detail, a significant contribution, especially when it comes to using the tomographic method, and that is the photo-z estimation.

Chapter 3

Photometric Redshifts

3.1 A galaxy's flux

3.1.1 luminosity function

As stated in the previous chapter, weak lensing focuses on determining matter distributions based on the statistical correlations on the distortions of galaxy images. As all the matter between the observer and the source affect the observed image, a good understanding of how far away the source images are (or, more precisely, the source images' redshift) is paramount. Determining redshift has already been covered previously in terms of standard candles but to record the redshifts of the millions of galaxies required for weak lensing surveys due to their low signal-to-noise ratio requires an entirely different technique. One based on astronomical photometry. This revolves around measuring the photon flux of galaxies over particular wavelength intervals. Intervals may reveal specific characteristics linked to redshift.

To consider a galaxy flux, we must first consider its luminosity and how it is linked to galactic variables. Luminosity is measured in units of $[J\,s^{-1}]$, revealing exactly what it measures; the amount of energy released by the galaxy every second. The luminosity of galaxies does not seem entirely random. Instead, it seems to follow a quantifiable distribution known as the galaxy luminosity function. At low luminosity, the number of galaxies rises with luminosity following a power law, but this number decays at high luminosity. Both the power-law rise and exponential plateau can be combined into a single function which states the number of galaxies that have a luminosity above L_{lim} (Schechter et al., 1976) :

$$N(L_{\text{lim}}) = \int_{L_{\text{lim}}}^{\infty} \phi_{\star} x^{\alpha} e^{-x} dx, \tag{3.1}$$

where ϕ is the normalisation parameter, and x is the ratio between the luminosity in the integral and the characteristic luminosity $L_\star$where the function transitions from power law to exponential. Thus, in this case, the parameters of interest are $\alpha, L_\star, \phi_\star$, which are determined from observation and fitted with data. The normalisation can cause issues, depending on how it is set up (does it normalise to the total number of galaxies or unity?). The luminosity function gives a number density instead of a number. This has led to the function being treated as PDF, which causes conflict with the fact that many known luminosity functions have been known to diverge at high luminosity (Dai et al., 2009). A more rigorous definition developed by Lake (2017) can be deployed, which states that given a luminosity interval dL_i around a luminosity L_i and a volume dV_i around a point in space x_i, the average number of sources that live in that luminosity-space interval is given as

$$< N_i >= \phi(L_i, x_i) dL_i dV_i. \tag{3.2}$$

Much like the averages we came across earlier with power spectra, the average is taken over all possible universes with different initial conditions, and N_i is the number of galaxies in that space-luminosity interval in one of those universes. However, as with previous power spectra, the homogeneity and isotropy of the universe ϕ are not dependent on x_i; thus, a spacial average can be used instead. Through this definition, one can assume that N_i is Poisson distributed and, through appropriate steps, form estimators based on observations. The parameters mentioned before are not necessarily permanent but can change with time and be different depending on the type of galaxy (spiral, elliptical, irregular). This particular dependency on time (redshift) and galaxy type comes into play in the templates below and is regularly used to generate realistic simulated galaxy samples.

3.1.2 Fluxes and magnitudes

In usual cases, we do not observe a galaxy's luminosity directly as this would require us to record all energy emitted by the galaxy in all directions. Instead, observing and measuring a galaxy's flux is more beneficial. This is the total energy of photons observed from a source per unit of area and time. One usually detects galactic electromagnetic radiation through a filter to gather more necessary information. This allows us to see the flux exclusively for specific wavelengths of light. The range of wavelengths filtered (otherwise known as filter width) are classified by photometric systems. Ranges above $> 300\,\text{Å}$ are considered broadband or wideband, in the range of $100\,\text{Å} - 300\,\text{Å}$ intermediate band and $< 100\,\text{Å}$ narrow band. different banding has its advantages. Telescopes are more sensitive to larger bands allowing fainter objects to be detected, while narrow bands allow for more detailed information.

If we consider a calculation, a filter will introduce a function $R_{\text{filter}}(\lambda)$ which states the probability of a photon of wavelength λ passing through the filter. For example, a perfect filter would be a top-hat function over relevant wavelengths. The flux would then be all the flux from that galaxy integrated over this function,

$$F = \int_0^\infty f(\lambda) R_{\text{filter}}(\lambda) d\lambda, \tag{3.3}$$

where $f(\lambda)$ is the galaxy's spectral energy distribution (SED). Measured in units of flux per unit wavelength, the observed SED is a characteristic of a galaxy that, much like the luminosity function, depends on redshift and galaxy type. The filter function (also known as the filter response) can also be represented as the product of 3 individual terms, all representing transmission curves as a function of wavelength. Transmission curves due to quantum efficiency of charge-coupled devices (T_{CCD}), curves due to the telescope's optics (T_{opt}) and curves due to atmospheric observatory conditions. The last curve is only important when observing from the earth. In space, this term has negligible effect, i.e. $T_{\text{atm}} \approx 1$.

Unfortunately, for historical reasons, the flux is not the standard unit used for observation. Instead, one usually relies on magnitudes, a logarithmic conversion of fluxes. Magnitudes are usually split between absolute and apparent. The apparent magnitude m is the brightness of an object as it appears on earth and is described as

$$m = -2.5 \log \frac{F}{F_0}, \tag{3.4}$$

where F is the flux, and F_0 is a normalising constant. This constant is also known as a zero-point flux, and itself can also be shown as a zero-point magnitude m_0 so that the apparent magnitude is also $m = -2.5 \log(F) + m_0$. Different magnitude systems use different zero-point magnitudes. For example, the Vega magnitude system sets the zero-point magnitude so that the star Vega would have 0 apparent magnitudes in all bands (Bessell, 2005) (as the star will be brighter in different bands, this means that m_0 can be band dependent). A more commonly used system is the AB system which uses a single reference flux of $F_0 = 3631\,Jy$ (Oke et al., 1983) or $m_0 = 48.6$ when measured in ergs^{-1} cm^{-2} Hz^{-1}. This magnitude corresponds to the flux density of Vega at about 5500 Å. Due to its single reference magnitude, the AB system makes comparing bands for a single object easy but can lead to difficulties when considering low signal to noise objects which can have negative fluxes due to noise.

An alternative and potentially superior system is that used by the SDSS survey and is known as the 'asinh magnitude system', which has the following equation (Lupton et al.,

1999):

$$m = -\frac{2.5}{\ln 10}\left[\sinh\frac{F}{2bF_0}^{-1} + \ln b\right], \tag{3.5}$$

where the F_0 parameter is the zero-point flux, the asinh magnitude system is almost identical to the AB system at the high signal-to-noise ratios. However, they differ at low signal to noise because of the b, the softening parameter defined as the 1σ noise of the sky in a point spread function aperture in 1" seeing. This addition essentially accounts for the noise in the measurement, allowing the magnitude system to behave reasonably well at low signal to noise even in the worst cases of negative fluxes caused by a large amount of noise. In both situations, AB magnitudes cannot.

As can be gathered from the name 'apparent magnitude', this is how bright the object is from our observation. This is not, however, a measure of the actual brightness of said object since observed fluxes will fall with distance squared. With this in mind, 2 apparent magnitudes of 2 different objects cannot be directly compared. Instead, one may wish to use the absolute magnitude, which is the apparent magnitude of the object *if* the object was at a fixed distance of 10 pc away from our observations. This is achieved through the following equation:

$$M = m - 5\log_{10}\left(\frac{D}{10\text{pc}}\right), \tag{3.6}$$

where M, m and D are the absolute magnitudes, apparent magnitudes and the distance the object is from us, respectively. To maximise information, the magnitude of a luminous object may be observed over a broad band of wavelengths. This brings with it issues, especially when observing multiple objects. If we presume that the objects are all not at the same redshift from us, then their fluxes have also been redshifted by different amounts, which means that if we were to use equation 3.6 to bring all objects to the same distance, it would still not be like-for-like as they would have been emitting different wavelengths of light to be detected by our broadband filter. With this in mind, a k-correction K_{QR} allows us to make the necessary correction in value to convert the apparent magnitude in broadband R to the absolute magnitude with rest frame broadband Q,

$$M_Q = m_R - 5\log_{10}(\frac{D_L}{10\text{pc}}) - K_{QR}, \tag{3.7}$$

where the distance D has been replaced with the luminosity distance D_L. This allows us to account for the fact that depending on the geometry and expansion of the universe, the relationships between 'distance' and 'flux red shift' would no longer be linear. The correction

itself can be estimated through the following equation (Hogg et al., 2002):

$$K = -2.5 \log \left[\frac{\int_0^\infty f_\lambda(\frac{\lambda}{1+z}) d\lambda}{\int_0^\infty f_\lambda(\lambda) R(\lambda) \lambda d\lambda} \right]. \tag{3.8}$$

The bottom of the equation gives the flux expected from an object in the apparent filter as seen in equation 3.3 where $R(\lambda)$ and $f(\lambda)$ are the filter response and flux density. While the top equation reveals a z redshift dependency. This can be incredibly useful for simulating the expected absolute magnitudes of galaxies at certain bands and distances (see below). With that said, it is not always required to use the k-correction. For example, if one uses an apparent magnitude using a bolometric system, there is no need for the correction. This is because the bolometric magnitude is based on light measured at all wavelengths instead of filters. This means the redshift in wavelength makes no difference to the result as all wavelengths are checked regardless. Secondly, if one is measuring known emission lines (see below), these lines will stay even if their position changes. Although not necessary for this thesis (the following is more useful for determining luminosity functions and stellar mass functions of galaxies), when making like-for-like comparisons of galaxies, one may not only want to correct the effect redshift has in that filter but also correct for the fact that you are observing galaxies at different points in their stellar evolution which can also affect the observed apparent magnitude. With this in mind, one can use an E-correction (Cole et al., 2001) which is usually modelled as linearly proportional to redshift so that one can add an (Ez) term to the end of equation 3.7. An example of this proportionality can be seen in the SDSS MGS sample which found E to be $1.07\,\mathrm{mag}\,z^{-1}$ (Saulder et al., 2013).

A correction we depend upon for weak lensing studies is the idea of interstellar (also known as galactic) extinction. Up to now, we have assumed the detectors detect every photon and wavelength in the direct line of observation. This, unfortunately, is not the case. Between object and observer, there is the interstellar medium consisting of dust and gas that will scatter and absorb the light passing through. To make matters worse, this cannot be modelled as a simple constant decrease in flux across all bands as the level of scatter/absorption is wavelength-dependent. For example, blue light scatters more than red light, leading to an artificially larger ratio of red to blue photons; galaxies are said to be 'dust reddened. With this in mind, different bands b will have a different amount of extinction A_b that is subtracted from the apparent magnitude m_b. To measure extinction, we usually use the term colour excess E_{B-V}, which states how much the B-V colour (colours are ratios between different bands) has reduced between intrinsic and observation.

$$\begin{aligned} E_{B-V} &= (B-V)_{\text{obs}} - (B-V)_{\text{int}} \\ &= A_B - A_V . \end{aligned} \tag{3.9}$$

Extinction is also direction-dependent as there will be a different amount of dust at different points in the sky; thus, the need for extinction maps showing their different extinction values are different points in the sky, such as the ones provided by (Schlegel et al., 1997) (or more recently re-calibrated by (Schlafly and Finkbeiner, 2010)). Finally, to make the final jump from colour excess to extinction, one needs to multiply E_{B-V} by the extinction curve

$$R_V = \frac{A_V}{E_{B-V}}. \tag{3.10}$$

Although R_V is also galaxy and line-of-sight dependent, we are mainly interested in earth-bound observations of which the commonly used value is $R_V = 3.1$. Every brand will have a slightly different value, but all can be multiplied by E_{B-V} to get the extinction in that band leading to large lists of known extinctions (Schlafly and Finkbeiner, 2010).

3.1.3 Spectroscopy

As seen above, a considerable effort is made to ensure that the light detected by our surveys is a true reflection of the light emitted by those galaxies. The reason behind such an effort being made is that due to the redshift effect of shifting the observed light to larger wavelengths, if we knew what the fluxes should be for any galaxies at particular wavelengths, then we can compare this to what is observed, revealing the redshift of the said galaxy.

Luckily enough, such calculations can be made. By looking at the fluxes across multiple very small binned wavelengths, one observes the spectral energy distribution of the galaxy (SED). These SEDs have common characteristics in their shape, usually in the form of sharp spikes and dips in recorded flux. These sharp spikes and dips (so sharp that they resemble straight lines) are known as emission and absorption lines, respectively. Figure 3.1 shows an example of a characteristic SED. This is the rest-spectra of a large red galaxy (LGR) from the SDSS Legacy survey, which has emission and absorption lines.

One may also notice that said absorption/emission lines have been labelled with corresponding elements. This is due to the nature in which such lines appear. Take emission lines, for example, taking place in star-forming galaxy regions where electrons are moving down an energy state, the difference in energy between the 2 states is directly proportional to the frequency of the emitted photon detected in the SED $E_\gamma = hf$ where E_γ is the energy of the emitted photon (the difference in energy states) and h and f are the plank constant and

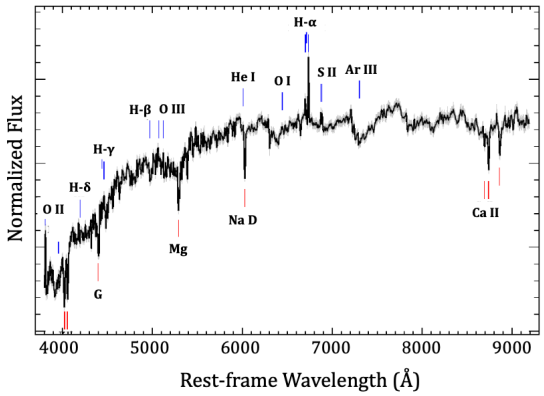


Figure 3.1: A rest-frame spectrum of an LRG obtained from the SDSS Legacy Survey, with the respective emission lines labelled SDSS-III collaboration et al. (2011).

frequency respectively. The size of these energy state differences is entirely dependent on the element of which the electron is occupying. Similarly are absorption lines. If certain elements are prominent within a galaxy, such as in gas, photons with frequencies matching energy states within these elements will be absorbed, leading to a dip in fluxes in the recorded SEDs. With this in mind, there are certain elements more prominent as well.

One of the most prominent (and, in fact, the most abundant) elements in our galaxies is hydrogen. Above, we have mentioned the *Lyman* $-\alpha$ line at a wavelength of 1215.7 Å, which is a consequence of electrons' transition from energy levels 2-1. As mentioned earlier, this absorption line is usually detected in Lyman-α forests and is not the usual approach for spectroscopy. Instead, the Balmer lines are the most recognisable characteristics of hydrogen within the SED. The Balmer lines are a series of emission lines (H-α, H-β, H-γ, H-δ, etc.) that consist of hydrogen electrons moving down to level 2. Starting with H-α, which is a 3-2 level drop appearing at the visible wavelength of 6563 Å, being visible the H-α emission line has direct consequences, giving Orion Nebula its pinkish look.

The abundance of hydrogen also has another consequence in the case of the Balmer break. This occurs in galaxies containing stars that are sufficiently hot enough that hydrogen atoms are ionising instead of going up energy levels. In this state, photons below a wavelength of 3646Å are continually absorbed by the free electrons leading to a vast lack of radiation below that point, as can be seen in Figure 3.2 (Massarotti et al., 2001). Furthermore, as stars have to be sufficiently hot, the break is usually linked to star-forming galaxies allowing these high redshifts 'Balmer-break' galaxies to be easily identifiable (Díaz Tello et al., 2016).

The last important hydrogen feature is the 21 cm line. Also known as the hydrogen line. Neutral hydrogen has a very fine energy split near the ground state between less energetic anti-parallel ground state (consisting of an electron and proton with a z-direction (m) spin of

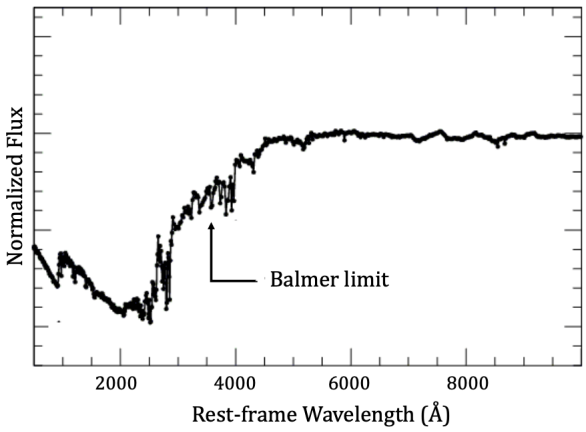


Figure 3.2: Rest-frame portion of galaxy SED explored by the WFPC2 photometric system at z = 0 and z = 2. The drop in flux represents the Balmer break (Massarotti et al., 2001).

$1/2$ and $-1/2$ respectively) and a more energetic parallel spin (with electron and proton spin of $1/2$ and $1/2$ respectively) state (Liu and Chronopoulos, 2008). As the electron transitions between the 2 fine states, a photon of wavelength 21 cm is released. Being this size makes this photon a radio wave which has the benefit of little interference between emitter and observer. Radio waves can pass through large clouds of interstellar dust unperturbed (in contrast, the dust would be opaque to visible light requiring accurate dust maps, as mentioned before). The main cosmological use (as opposed to the main use overall which lies with radio astronomy, such as the mapping of galactic spiral structure) of 21 cm lines today is in probing the universe between recombination and reionisation, as ionised hydrogen will not have the expected 21 cm characteristic. Unfortunately, due to its radio wave nature, 21 cm lines do not make the best candidates for large ground-based surveys due to their interference with television transmitters and the ionosphere; thus, only very secluded locations, such as the far side of the moon (Burns et al., 2021) can provide the kind of required data. As demonstrated above, there are very distinct features in galaxy SED (more could have been mentioned, such as the Lyman Break at 1214 Å). Due to this distinctiveness, the redshifts determined from spectroscopy can be expected to have a precision better than 10^{-3} (Salvato et al., 2018). As good as that may be, spectroscopy is not the main redshift determiner of large weak lensing surveys. This has a lot to do with the level of precision needed in the probing wavelengths, usually requiring a resolving power of $R = \lambda/\delta\lambda > 200$. Resolving large power requires significant time to absorb photons in each small wavelength bin for sufficient signal to noise. Even with the efficiency of today's multi-object spectrographs (i.e. VLT/VIMOS, KECK/DEIMOS, SUBARU/FMOS), meaningful spectra can only be obtained for a tiny percentage of detected sources in deep imaging surveys. Even among

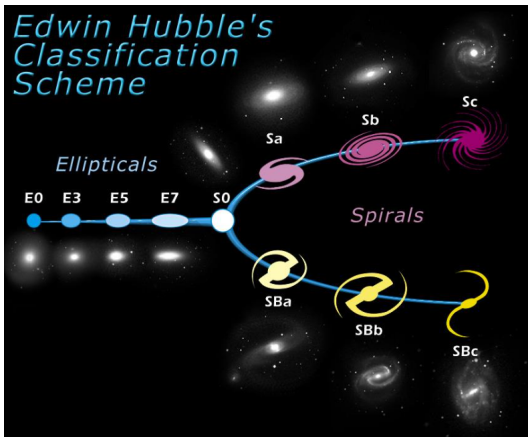


Figure 3.3: An illustration of the Edwin Hubble classification system and how it relates to galaxy shape and morphology (NASA & ESA, 1999).

this small percentage, a minimum of 2 well-defined spectral features (the kind mentioned above) are required for a precise redshift measurement. Given the limited signal-to-noise and spectral coverage of some deep surveys, this cannot be easy to achieve (Salvato et al., 2018). Especially in the case of weak lensing surveys where one is looking at very faint objects, the success rate of measuring a spectroscopic redshift (remembering that this is already from a small percentage subset of the full survey) can be as low as $50\% - 70\%$ (Newman et al., 2013a). This is not an applicable method going forward.

3.1.4 Morphology

Although spectroscopy cannot be used directly, it offers clues as to the correct method to scan extensive surveys and determine their redshifts efficiently. As stated, galaxies have standard spectral features, but as hinted by the Balmer-Break galaxies, some will show particular features more than others, and thus not all SEDs will look the same. A better understanding of what determines certain features may add insight and efficiency to using these features for redshift identification. One such determination is the galaxy morphology which refers to the structural features of the galaxy.

The Edwin Hubble Classification scheme is the universal system based on galaxy morphology. A scheme illustrated in Figure 3.3. As you can see from the illustration, in general, there are 2 galaxy shapes used; elliptical galaxies and spiral galaxies. As the name suggests, Elliptical galaxies have ellipsoidal shapes, and the level of brightness across the galaxy varies very little. Thus, the brightness profile appears featureless and smooth. Galaxies of this nature are classified by an 'E' followed by a number that increases with galaxy ellipticity. Much like the ellipticity numbers seen before, this number is a ratio of the major and minor

axis,

$$\text{classification number} = \lfloor 10(\frac{b}{a}) \rceil, \tag{3.11}$$

where b and a denote minor and major axis and $\lfloor x \rceil$ means rounding x to the nearest integer. Elliptical galaxies are theorised to be the oldest of the galaxies, being from early in the universe, they have mainly older low mass stars and be redder in appearance. Their smoothness and elliptical nature may be directly linked to this extra time, and years of galactic conflict (as in collisions and mergers with nearby galaxies) have robbed them of their original shape and outermost stars. Although such major galactic mergers (major referring to the galaxies involved being of comparable mass) were more common in the early universe, it has to be theorised that an elliptical galaxy will be the result of the collision between the Andromeda galaxy and our very own Milky Way galaxy in about 4-5 billion years (Nagamine and Loeb, 2002).

In contrast, spiral galaxies have a prominent disk-like structure and are usually complemented with bars, spiral shapes and bulges. Within the classification system, spiral galaxies are named with an 'S' followed by 'B' if the galaxy has a bar (which is believed to be roughly two-thirds of all spirals (Mihalas and Routly, 1968)) or 'a,b, or c' if it does not. Unlike the elliptical galaxies, they are usually bluer in colour, full of young active stars' ongoing star formation, and due to different formation levels, the brightness profile has unsmoothed features. The Milky Way and our neighbour, the Andromeda galaxy, are spiral. When it comes to how the spiral shape exists in the first place, problems have occurred. The spiral shape seems to be one that cannot co-exist with the idea that galaxies exhibit differential rotation. Stars within the disk have the same orbital speed so that stars further along the arms with longer orbits will lag, tightening the spiral arms until they ultimately disappear. This is known as the winding problem and has led to the 2 leading ideas of how spiral arms are possible being the density wave model (the arms are the consequence of regions of the galaxy with higher densities (Davis et al., 2003)) and the stochastic self-propagating star formation model which says the arms formed due to shock waves in the interstellar medium (Mueller et al., 1976).

There are 2 other main classes of the galaxy we must consider. The first is lenticular galaxies. In Figure 3.3 they are classified as 'S0' and represent the in-between of spirals and ellipticals. They still have a 2-dimensional disk-like structure seen in spiral galaxies (unlike the more 3-d ellipticals) but do not have large spiral arms. They are believed to be disk galaxies that have used up or lost most of their interstellar mass; this means a minimal level of star formation is taking place, so the majority of the stars within it are ageing (like ellipticals). Sub-classifications can continue, such as 'S0' or 'SB0' if a bar is present, subscripts like '$S0_1$' referring to dust absorption levels in the disk component or 'ES' galaxies classifying those

that the sub-in-between of ellipticals to lenticulars. The other relevant class are the irregulars which refer to galaxies that do not follow the elliptical and spiral-based structure and have no distinct regular shape. Although their lack of regularity suggests rareness, they make up a quarter of all known galaxies. This commonness has led to some suggesting irregulars used to be spirals and ellipticals that have been deformed by external gravitational effects such as collision and minor mergers with much larger galaxies. This idea has additional weight as irregulars are usually small and about a tenth of the size of the Milky Way. Just like lenticular galaxies. Irregulars also have subclasses (Irr I, Irr II and dIrr) which refer to the galaxy having some spiral/elliptical structure, no structure at all or being a dwarf in size.

The simplicity of the Hubble classification has been criticised regardless of its everyday use leading to calls for updating its structure (Masters et al., 2019). With that said, it is common to see galaxies classified in an extended form of the Hubble system known as the de Vaucouleurs system (de Vaucouleurs et al., 1959), which follows the Hubble system with a few changes.

- non-barred spirals are now denoted as 'SA' as opposed to just S with lenticulars also split based on having a bar or not (SA0, SB0)
- a third spiral fork is added, which is denoted by 'SAB' as in intermediate spiral has a weak bar
- many Ir I galaxies have been reclassified as spirals, such as 'Sd', which are spirals with diffused broken arms, and 'Sm', which are spirals without a central bulge
- galaxies are also denoted by (r),(s) or (rs) based on if they have had, have not or are in a transition stage of a ring-like structure.

Combined, this leads to a much more detailed morphology structure, such as a non-barred spiral galaxy with broken arms and no rings classified as 'SA(s)d'.

On the surface, this may all seem like pointless taxonomy, but this is very relevant to identifying identifiable SED structures. As stated earlier, not all galaxies have the same rest frame SED and looking for common structures may be inefficient and inaccurate. So instead, we can use their taxonomy to narrow down our search (in contrast, widen our parameter space) in pursuit of more accuracy. Firstly we consider the typical spectra of elliptical galaxies. As stated, elliptical galaxies are usually populated by early-type galaxies and thus contain older, cooler giants. The typical SED of an elliptical galaxy can be seen in Figure 3.4 the SED is characterised by a few key features. Firstly an abundance of absorption lines. This is because most heavier elements have already been created, and thus, light is absorbed, leading to typical Magnesium, Sodium and Calcium lines at 5175 Å,5894 Å, and

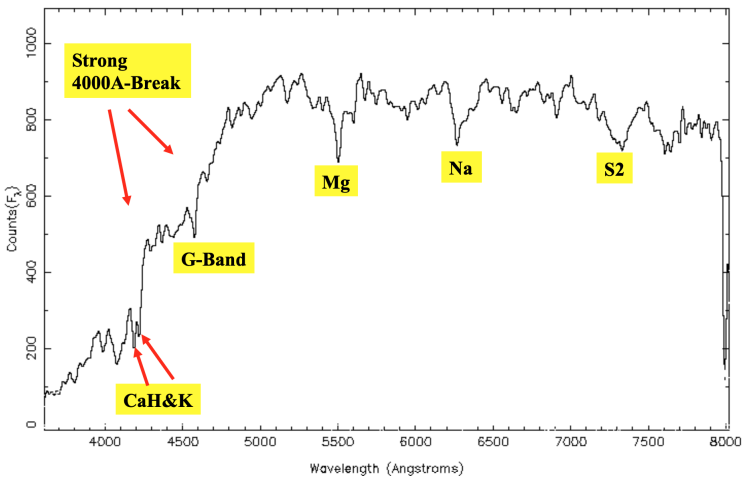


Figure 3.4: A typical SED from an elliptical galaxy showing strong absorption lines due to metals in the stellar atmospheres of a mostly low luminosity stellar population while no evidence of any emission lines and hence no young stars and no gas (Driver, 2014).

3969 Å respectively. In addition, due to the lack of young stars, there is not a vital blue/UV component to these galaxies allowing for a very clear Balmer Break. Another feature is the complete lack of emission lines due to them not having H II regions (interstellar matter consisting of ionised Hydrogen). Similar spectra can also be seen among the lenticulars (S0).

In contrast, the SEDs of spiral galaxies (see Figure 3.5) contain emission lines, especially the solid H-α emission line in the H II region heated by hot young stars. A small amount of absorption from underlying older stellar populations. Moreover, a much less defined Balmer break is due to the presence of bluer hot stars. Even more extreme are the irregular spectra (see Figure 3.6). Having very little real structure, the galaxies are mainly entirely young, starting with a large H II gas region, thus no Balmer break at all, no absorption lines and extensive emission lines along the Balmer series.

As these features seem to be dependent on the type of galaxy we are looking at, more precise ideas of what their redshift can be explored as long as these 'types' can be parameterised. We only expect certain features from certain galaxies that allow for a total redshift estimation. For example, the 4000 Å break mentioned above is a clear indicator of an LRG (luminous red galaxy), which is most likely to also be elliptical for the reasons explored above. Due to the broadness of the break, identification of an LRG would not require complete spectroscopy, but if said identification was made in the optical range, then a redshift estimate of $0.12 < z < 0.55$ is already possible. Absorption lines are also associated mainly with nearby low-redshift galaxies as less than 10% of these galaxies have strong emission lines (Bolton et al., 2012). Bluer galaxies, on the other hand, are usually classified as ELGs (emission-line galaxies) where the detection of just OIII lines in the optical range allows for a quick redshift

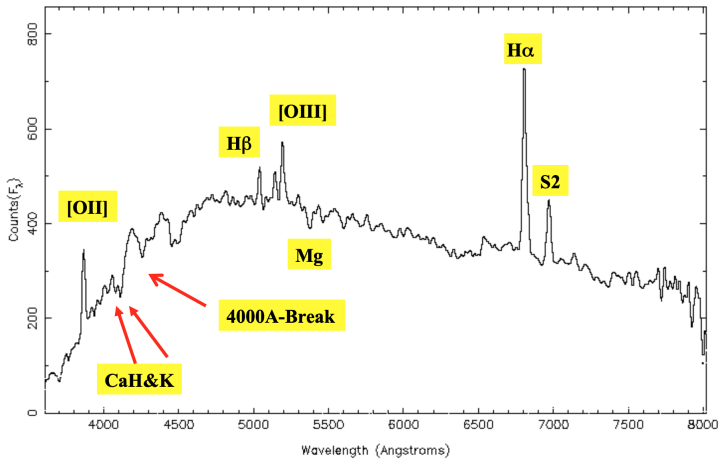


Figure 3.5: A typical SED from a spiral galaxy. Showing some emission and some absorption indicating both a young and old stellar population Driver (2014).

estimation of 0.25<z<0.68. Other emission lines are also primarily associated with another classification of the galaxy; the AGN (active galactic nuclei). These galaxies currently have a supermassive galaxy at their core that continuously releases matter, leading to a much higher density and luminosity at their centre. Due to the high speed at which matter is released from the supermassive black hole, photons from that matter that are released will be Doppler shifted (the changing observed wavelength due to the motion of the emitter) in both the blue and red direction as there would be no consistency in the motion of the matter relative to us. For example, the matter consisted of hydrogen atoms releasing $H-\alpha$ emission lines. If the velocity of the matter relative to us were in a range of Δv, then the range of wavelengths, all of which cover the $H-\alpha$ emission lines, would be $\Delta\lambda = \lambda_{H_\alpha}\Delta v/c$ where λ_{H_α} is the rest-frame wavelength of the $H-\alpha$ emission line. This spread means the emission does not appear as a strict line but instead is a more broad peak. This effect is known as Doppler Broadening (Kollatschny and Zetzl, 2013). This extends to quasars (a specific form of AGN) with bright blue continuum and broad lines, prominent non-line-based features that allow for redshift estimation beyond the redshift desert, a region at $1.5 < z < 2.5$ where optical and ultraviolet observations in the past have been complex.

As stated previously, measuring individual wavelengths (and thus spectroscopy) is not a viable option for the number of galaxy redshifts required for a weak lensing survey. Nevertheless, the added information that not all galaxy types have identical expected spectra and that these different types may have much broader SED features than single absorption and emission lines open the door to a different approach that still uses galaxy SED but in a much more efficient manner.

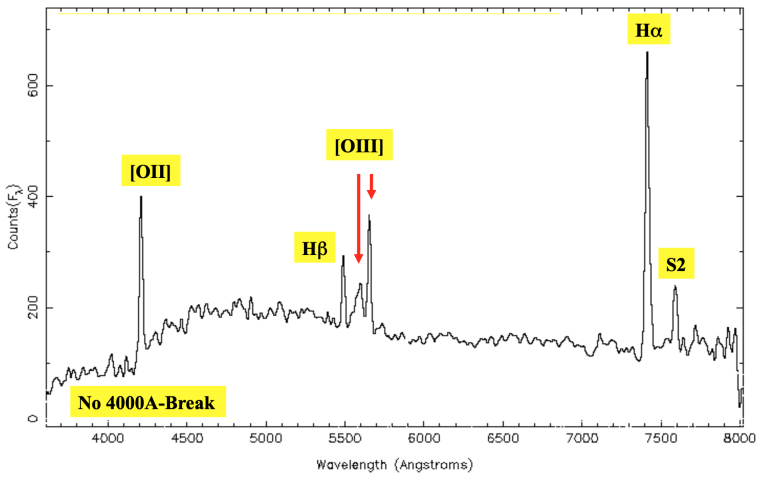


Figure 3.6: A typical SED from an irregular galaxy. A strong emission-line spectrum indicating many hot young stars heating the gas which is re-radiating at specific wavelengths which depend on the chemical composition of the gas (Driver, 2014).

3.2 Photometry

3.2.1 Origin

Although the above may sound trivial. The main principle of photometry and thus estimating the photometric redshift (otherwise known as photo-z) is to take the same approach seen in spectroscopy (look for SED features of known rest-frame wavelength and use the difference in observed wavelength to determine redshift) but with much broader filters. To indicate how broad, the spectral resolution of spectroscopy was around $R > 200$ and the spectral resolution of photometry was at least 4 times less. So much so that photo-zs were defined as a Redshift estimation using only images of spectral resolution $R < 20$ (Koo, 1999). I say 'at one point' as modern photometric surveys have continued to blur the line between 'photo-z' and 'spec-z'. So, for example, the redshifts obtained by PRIMUS (Prism Multi-object survey) have a spectral resolution of $R \sim 40$ (Coil et al., 2011) are considered spectroscopy. In contrast, the redshifts obtained from PAUS (Physics of Accelerating universe survey) have a resolution of $R \sim 50$ and are considered photometric (Cabayol et al., 2021). This is an exceptional case where the survey used a 40 narrow-band filter camera (typical photometric surveys have 3-10 broadband filters) which still has fewer data points than spectroscopy's wavelength to wavelength information.

The first known use of photometric redshifts was by Baum (1962). Where a photoelectric photometer (an instrument that measures the brightness of stars through electric currents produced when photons interact with the light-sensitive surface, a consequence of the

photoelectric effect) constrained by just 9 photometric filters was able to span the spectra from 3730 Å to 9875 Å. Using this system, he took the 9 filter SED of elliptical galaxies in the Virgo cluster and elliptical galaxies in the Cl0925-2044 (also known as Abell 0821). The redshifts of both clusters were spectroscopically known, but this was a demonstration that by only knowing the redshift of one (in this case, Virgo), you could estimate the other. This was achieved by plotting the average SED of both clusters on a log wavelength scale to discover the displacement in the energy distributions. Through this technique, he was able to estimate a redshift for Abell 0821 of $z_p = 0.19$ (very close to the spec-z estimate of $z_{spec} = 0.192$). The method was so effective that he could extend the approach to galaxies at much higher unknown redshifts of $z_p = 0.46$. However, with this effectiveness came a severe limitation. Such broad comparisons between galaxy SEDs require them to have very broad features and similar features to compare; thus, the method was limited to the broad feature of the 4000Å break, which is only identifiable in elliptical galaxies, as seen above. With that said, he did find other uses for his photometry, such as in estimating differences in stellar populations of spiral galaxy arms (Baum, 1966).

Another early candidate for using photometry to estimate photometric redshift came from examining the broad U band. Suppose we have a situation where an otherwise very blue galaxy and a significant drop in its bluest band, then it's an indication that the U band has entered the Lyman break, a phenomenon that occurs at $z \sim 3$ and was known as the "Lyman-break technique" (Partridge et al., 1974). Although technically, this same drop can also be achieved without the Lyman breakthrough absorption in the Lyman-alpha forest, it should be known as the "Lyman-drop technique" (Koo, 1999).

In both these cases, the phrase 'photometric redshift' was never said. The first reference to photometric redshift did not come till the work of Puschell et al. (1982). Although the only reference was in the abstract, their attempt to estimate redshifts through photometry of faint galaxies set up many of the modern standards approaches to photometric redshifts, such as the use of near-infrared bands (*JHK*) alongside optical bands (*RI*), the use of SED templates ranging across many different types of galaxies, whose origins are both organic and synthetic as well as the match of SEDs to templates through a χ^2 method. These are still methods at the heart of modern-day template-based photometry (see template-based methods below). When combined with the technological advancements introduced by Koo et al. (1985) where he used photographic plates (glass plates coated by a light-sensitive emulsion of silver salts, made popular through their ability to not shrink or deform under environmental changes common to astronomical study) to measure the photometric redshifts of multiple galaxies simultaneously and with a relatively small amount of filters (U, B_J, R_F, I_N), he showed that with ease (relative ease compared to spectroscopy). He could estimate the redshifts of over

100 galaxy samples. These building blocks led to photometric redshifts being the core of many multiwavelength surveys today. Due to their lack of time and input data needed per galaxy, fainter galaxies (and thus deeper surveys) were possible. These estimations would only grow in confidence and validation with every new spec-z survey. The fact that some of the most state-of-the-art photo-z codes are available today has contributed to publications, including the term 'photometric redshift' increasing by a factor of 10 within the last 2 decades (Salvato et al., 2018).

In the modern era of photometric redshifts, the methodology fills sharply into 2 different camps: template-based (otherwise known as physically motivated) or empirical-based (also known as data-driven). Both camps have their distinct advantages and disadvantages when compared to each other. For example, empirical methods are considered more likely to give an accurate photo-z estimate so that when compared and validated against their spec-z counterparts (where the spec-z redshift is considered correct), they are less likely to produce outliers than template-based methods (Hildebrandt et al., 2010). However, if we were to eradicate outliers from the analysis, we can use a metric of validation known as the *68th percentile error* σ_{68}

$$\sigma_{68} = \frac{Q_{84.1\%}(\Delta z_i) - Q_{15.9\%}(\Delta z_i)}{2}, \tag{3.12}$$

where $\Delta z_i = (z_{\text{phot},i} - z_{\text{spec},i})/(1 + z_{\text{spec},i})$ measures the error between the perceived correct spectroscopic redshift z_{spec} and the the estimated photometric redshift z_{phot} scaled by $1 + z_{\text{spec},i}$. Q are the quantifies of the distribution so that σ_{68} only considers the error around the core meaning of the photo-z distribution. Furthermore, empirical methods perform better at estimating intermediate redshifts, whereas template methods perform better at the extremes (Abdalla et al., 2011). The reason for this can be explained through their alternative names and indicates how both methods are used side by side today. Empirical methods are data-driven, so they can perform optimally in intermediate redshifts in which much spec-z data has been collected. However, the results are expectantly worse at redshift extremes, where there are less data to 'train' on. Template methods, i.e. Physically motivated methods will not necessarily have the same issue, relying on physical assumptions that should hold regardless of the amount of spec-z data collected in the region. Given that their advantages directly affect different redshift regions, attempts have been made to integrate them into each other (Wolf, 2009) or to combine the estimations from both methods cohesively (Dahlen et al., 2013). Both methods can be considered combined because, fundamentally, they are doing the same things to achieve the same goal. Consider the flow diagram in Figure 3.7 which shows fundamental steps to any photo-z estimation regardless of camp. All photo-z techniques manipulate the key features we are used to seeing in SEDs and isolate the redshifted wavelength of which these features occur compared to at rest frame. The flow

diagram shows that all the sources may change depending on the method (be it spectroscopic samples or physical models). The first step is to create a model mapping source fluxes (or colours which are the ratio of different filter bands) to redshift. One such mapping model can be seen within the 'colour redshift mapping model', where the grey points represent real galaxies on the redshift-colour plane while the solid lines represent the expected redshift-colour relationship based on these galaxies. The model is then compared to the input data, which in this case would be a series of flux or colours in an average of 4-10 available filters. Optionally a preliminary (extra information outside of the model) may or may not be added. This comparison allows us to find either the single best redshift that reflects the data (a point estimate z_{phot}) or a series of probabilities of that data existing at a range of redshifts known as a probability distribution function $p(z_{\text{phot}})$ of PDF. At this point, results validation would occur directly through spectroscopic comparison or indirectly through other statistical methods such as cross-correlation. The importance and relevance of priors, probability distributions and validation techniques will be explored in more detail below. As stated, the flow diagram represents both camps of methodology, with the real difference being how one gets from step to step, such as how the redshift-flux mapping model is created (and the information used to create it) and the tools used to make the model-input data comparison. Although the work of this thesis is based on a template-based method, some of the ideas pushed forward can still be applied to empirical-based methods. Thus, a brief description of that method will be given below.

3.2.2 Empirical Methods

The Empirical method camp of photometric redshift estimation developed much later than template-based methods and, in one sense, was a direct response to some of the known criticisms of the said method, mainly the fact that template methods (as will be explored below) is dependent on physical models that may change with time. Thus Connolly et al. (1995) developed a model-independent approach that only required a sample of galaxies with photometric data and known spectroscopic redshifts. Using this data, he developed a polynomial relationship between redshift and magnitude, both linearly and more accurately quadratic:

$$z(m) = a_0 + \sum_{i=1}^{N} a_i m_i, \tag{3.13}$$

$$z(m) = a_0 + \sum_{i=1}^{N} a_i m_i + \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} m_i m_j, \tag{3.14}$$

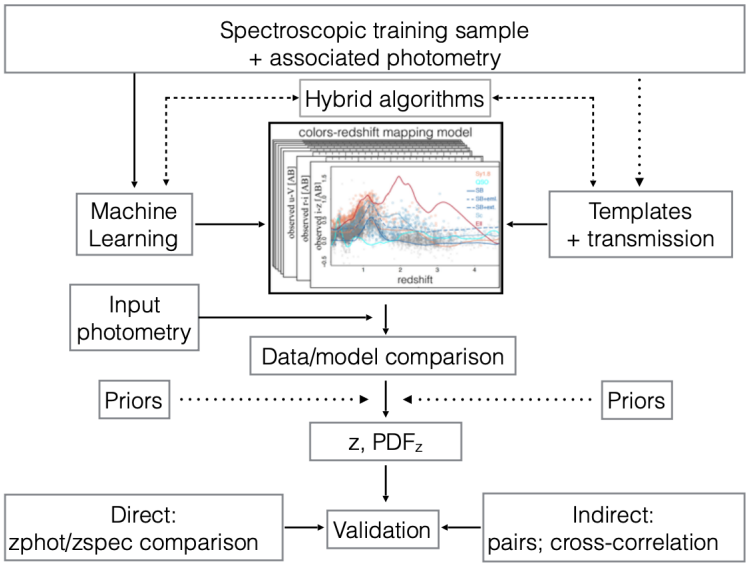


Figure 3.7: A flow diagram showing the fundamental steps required for any photo-z estimation, empirical or template-based methods. The dotted lines represent optional steps. The figure within the colour-redshift mapping model represents galaxies' spec-z results compared to their colours (Salvato et al., 2018).

where N is the number of filters, m_i is the magnitude at one of these filters and $a_0 a_i, a_j$ are the relevant coefficients found through a linear regression (or least squares minimisation) technique. Using just 4 filters (a 4 bandpass system of *UJFN*) and 370 galaxies these models were able to determine redshifts up to $z = 0.5$ with uncertainties of $\sigma_z = 0.057$ and $\sigma_z = 0.047$ for the linear and quadratic models respectively. Further analysis showed that through this method, there was little to no loss of information if one used colours (flux ratios, i.e. $C_i = M_i - M_{i+1}$) instead of magnitudes increasing versatility. In this case, much like the earliest photometric work, the 4000Å break was the main contributor, so this first attempt may have been limiting in scope. With that said the sheer ease with which this method can be used (as one just needs spec-z data with no need for working theories of redshift and galaxy formation) lead to many quick uses, such as the measuring of the luminosity function up to $J = 24$ (SubbaRao et al., 1996). The reliance on the 4000Å was also mitigated, showing that it is not a vital and thus limiting factor for empirical approaches by Richards et al. (2001). Using a small sample of quasars of known redshift as well as 4 SDSS colours ($u-g, g-r, r-i, i-z$), a colour-redshift relation (CZR) is formed by plotting these quasar's colours against redshift and taking the mean colours across a set of redshift bins (through interpolation). A χ^2 function of redshift is then formed by comparing the observed colours

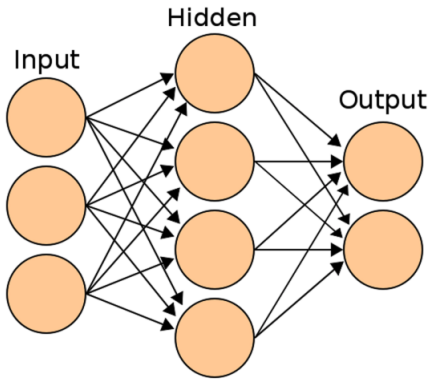


Figure 3.8: A simplified Neural Network consisting of an input, hidden and output layer (Bonnett, 2015).

to the CZR:

$$\chi^2 = \sum_i^N \frac{(C_{i,\mathrm{obs}} - C_{i,\mathrm{CZR}}(z))^2}{\sigma^2_{C_{i,\mathrm{obs}}} + \sigma^2_{C_{i,\mathrm{CZR}}}}, \tag{3.15}$$

where $C_{i,\mathrm{obs}}$ are the observed colours, $C_{i,\mathrm{CZR}}(z)$ the mean colours of the CZR at a given redshift, and σ the uncertainties on those colours, the redshift with the lowest χ^2 becomes the redshift estimate. This method was improved upon by Weinstein et al. (2004) by making 4 changes that are now also seen as hallmarks of the modern era of photometric estimation:

1. As opposed to linear interpolation, the CDZ, is parameterised in a more refined way using a series of 4-dimensional multivariate Gaussians where the variance of the gaussian (or the CDZ 'width') can change as a function of redshift

2. Uncertainties are handled more rigorously; colour uncertainties are not independent, meaning the quasar colour convenience is not 0. Therefore a full covariance matrix is considered in the χ^2 as opposed to individual uncertainties

3. As a consequence of the use of a covariance matrix and the varying width of the CDZ χ^2 is no longer a stand-in for probability, so instead of minimising χ^2, we look to maximise redshift probability $p(z)$ (this will be used in more detail below)

These improvements led to the result that 90% of quasars across the extensive redshift range of $0.8 < z_{phot} < 2.2$ can be estimated with an accuracy of $\Delta z < 0.3$. The last 2 refinements become exceptionally important mainly as we discuss Bayesian approaches below.

The most modern version of this data-driven approach moves strictly away from the analytical functions and polynomial models above and instead embraces machine learning as the next step in complete model independence (as, in a sense, by fitting coefficients to an analytical function, you are assuming redshift-flux mapping follows a straightforward model).

Much like photometric techniques in general, machine learning can also be divided into 2 camps; supervised and unsupervised. Supervised methods most mimic what we have seen previously. They require a set of training data consisting of galaxy sources with photo-zs and spec-zs for the method to 'learn' a function that maps one to the other. By 'learning', the algorithm finds the set of coefficients that best fit the data and interpolates much like the early empirical methods before it, but this time on a much larger dimensional scale and without the idea of analytically constructing the function of which it is based. This has its advantages. The constants on the possible mapping are way less, allowing for a much better and thus more accurate fitting, but on the other hand, these methods can act like 'black boxes' giving only inputs and outputs with no understanding of how one went from one to the other. To understand the vagueness and complexity of these supervised methods, consider the 2 most popular approaches:

1. Random Forests (Kind and Brunner, 2013a)(Carliles et al., 2010). Aptly called because it consists of a collection of 'decision trees' that have all been independently taught on a random subset of the training data. Each decision tree aims to split the data set in two along some axis of the input photo-z data. This splitting axis is determined by minimising the 'resubstitution error' $\varepsilon_{\text{resubs}} = \varepsilon_1 + \varepsilon_2$ where ε_i is the standard deviation away from the mean spectroscopic redshift of that split subset. The amount of splitting (or branching) continues until it no longer gives a performance benefit. What is left is a set of logical if conditions leading the input data to a small set of spec-zs of which the mean spec-z becomes the input assigned value. The fact that we repeat this across many trees directly counters the main disadvantages of single decision trees. Firstly they are prone to overfitting (accurate with data that looks the same as it was trained on but catastrophic outside of that data space), so the forced setting of the forest forces variance in the output value and that a single tree can only give values, not uncertainty. On the other hand, the forest takes the output values across every tree to give a mean and uncertainty photo-z estimate, a necessity for modern weak lensing surveys. The main disadvantage of this approach is the sheer time and memory consumption required to run data through the necessary amount of trees, and the fact that your taking an average of many large logical conditions completely removes any chance of interpretability of how one got their outcome. A study using a random forest consisting of 400 trees trained on 80000 objects and tested on 100,000 objects from the SDSS survey (Carliles et al., 2010) found an accuracy level comparable to other modern approaches. What is most important is that due to a lack of any underlying parametric model, the bias was essentially zero.

2. Neural Networks (Bonnett, 2015). Designed originally as an analogue to how humans and animals think, a set of inputs (the photometric data) interact with several layers of 'neurons' that interact with all the inputs based on weights and biases. These same neurons act as the input to the following layers until we have an output in the form of a redshift estimation. Let us consider that in slightly more detail. Figure 3.8 is a basic example of a Neural Network consisting of 3 input and 2 output nodes. Following the notation of Mackay (1995), the mathematics is the same whether we go from input to hidden or hidden to output. Either way, we refer to the values of all the prior layers nodes as x and the node of the following layer values as y then given there are i prior layer nodes and j following layer nodes and $k+1$ layers (the $+1$ is due to the input layer having no prior interaction) all together we get the following transformation:

$$y_j = g^{(k)}(f_i^{(k)}), \tag{3.16}$$

$$f_i^{(k)} = \theta_j^{(k)} + \sum_i w_{ij} x_i. \tag{3.17}$$

For consistency its worth noting that if x_i is the k^{th} layer the $y_j = x_j^{k+1}$. From equation 3.17 we can see a set of biases $\theta_j^{(k)}$ and weights w_{ij} that control the output variable. It is worth noting that these weights change for every layer, input node and output node, while the biases only change per layer and output. The only large mathematical change that can be seen is in equation 3.16 where $g^{(k)}$ is an interaction function, the decider of how the next layer interprets the interactions of the previous. In simple cases this can simply be $g^{(k)}(x) = x$ as in to say that the weights and biases are used directly but in more complicated cases it might be advised to use a sigmoid function $g^{(k)}(x) = 1/(1+e^{-x})$. One example would be if the output is a probability and thus must live within the $[0,1]$ range. Adding this function to the final layer would ensure this. The number of layers, number of nodes and use of particular $g^{(k)}(x)$ are all at the discretion of the user; however, when these are chosen, the network goes through a learning stage where a set of inputs and true outputs are known to optimise the weights and biases. For example, the above paper used a second-order optimisation method known as conjugate gradient algorithm (Graff et al., 2014). Because so many options to the neural network structure allow for great flexibility in use, take for example (Bonnett, 2015) which does not have a single output node representing the estimated redshift. Instead, there is a range of output nodes, each giving a probability value of the source within a certain redshift range. This approach produced a set of photometric estimations optimised for weak lensing study using a sample of 58714

galaxies in CFHTLenS that have spectroscopic redshifts from DEEP2, VVDS and VIPERS. Furthermore, the main weakness of Neural Networks, like most machine learning codes, is correctly estimating inputs outside of the training samples data space. Even with that said, the effect on the CFHTLensS results was a slight estimation of how many galaxies have high redshifts (only 0.2%-0.4%), an amount small enough that it had little effect on the final weak lensing results. Although the optimisation of the weights may seem close to the polymetric coefficients from above, their sheer quantity and the structure's flexibility are not beholden to an underlying model in the same way. The quantity of neurons (and thus weights and biases) has increased exponentially in recent years, leading to the current state of the art of modern computer science and photo-z estimation; Deep Machine Learning. These methods can be so sophisticated that work has taken place to make individual photometric data points, pre-classifications and known errors such as extraction completely obsolete by using the photometric *images themselves* the input variable (D'Isanto and Polsterer, 2017). This means extra information that may help inform redshift estimations such as galaxy morphology (see 3.1.4) or dust map corrections can be 'learnt' while never directly inputted.

In the other camp of machine learning, we have unsupervised learning. This approach goes even further in trying to achieve model independence by not involving redshift values in the model creation stage; thus, the tendency to overfit a model around the redshift values is wholly removed. To explain how this works in practice, one of the leading unsupervised machine learning photometric estimators is SOMz (photometric redshift PDFs with self-organising maps and random atlas) (Kind and Brunner, 2013b). Taking a similar approach to their previously mentioned work TPZ (Photometric redshift PDFs and ancillary information by using prediction trees and random forests) (Kind and Brunner, 2013a) both methods aim to have a set of cells consisting of similar galaxies and then estimate redshifts based on the cells mean redshift. The main difference between the 2 is that while TPZ organisation of the cells depends on their redshifts (as each cell splitting was aimed to reduce scatter), the SOMz method does not use redshift data at all (outside of a culling technique that will be elaborated on in further chapters). Instead, the cells are based on what galaxies are most similar in input values alone. Thus a number-of-filters dimensional data set is projected onto a 2-dimensional map. Only after the map is created are the spec-z values added to the map, as shown in Figure 3.9. One of the main weaknesses of this approach is that due to the lack of 'supervision', galaxies may be grouped with vastly different redshifts, and thus the map would be unusable. However, due to the apparent correlation between photometric data and redshift, the maps indicate a usable model with some cells having a much larger mean

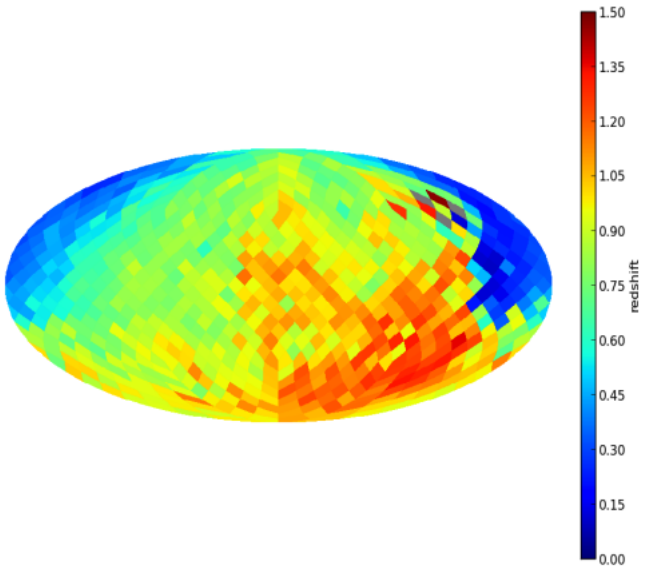


Figure 3.9: An example of a topological map consisting of 768 cells on a spherical HEALPix grid was created by SOMz of which the spec-z values have been added in post (Kind and Brunner, 2013b).

redshift than others. The creation of the map itself is beyond the scope of this thesis, but the procedure for estimating redshift is not. A new input source is placed in the cell of most similar galaxies, and the estimated redshift is that of the cell. Much like many random trees are used to create a random forest in TPZ, SOMz uses a random atlas, many 1000s of maps created to give uncertainties and reduce overfitting in the same way. When both models were directly compared, their accuracy was comparable, with the main advantage of SOMz being that the choices one would make beforehand (number of cells on a map, size of cells, shape of calls and map, number of maps, etc.) had much less of an effect than the choices made for TPZ (number of trees, number of branches, the final number of galaxies per cell, etc.) meaning a much less intrusive meta-optimisation stage.

No matter the individual approach, all empirical techniques share advantages and disadvantages compared to template-based methods. For starters, they vastly outperform the template counterparts when it comes to speed, mainly because, being quite a modern idea, they have been optimised for the large datasets commonly used today (Vanzella et al., 2004). Also, given the circumstances of being trained on a wealth of training data represented by your input data, empirical methods usually outperform template-based methods in accuracy. However, here lies the central ensure of data-driven methods, how much can be learnt? If the methods rely on a load of spec-z data, then the redshift distributions in these regions may already be known. Any input data that differs from the norm suffers from extrapolation of the model; thus, data-driven approaches suffer considerable losses in accuracy when

used in deep-pencil bean studies covering large redshift ranges (Hildebrandt et al., 2012). Moreover, it can be difficult to tell if this extrapolation has occurred as there would be such a small amount of spec-z data to which to compare it. Lastly, although the model Independence means the algorithms may come across correlations between specific data points and redshifts, those correlations remain unknown (one could analyse the weight of individual neurons to see what inputs they favour, but after being multiple layers deep, the results may be meaningless). This makes these methods harder to discover distributions of new regions and for new physics.

3.2.3 Template Based Methods

Template-based methods (or physically motivated) methods are the alternative approaches and are the older of the 2. Going forward, all methods and techniques will be about the template method, although some of the later techniques could potentially still be used within an empirical-based framework.

Template fitting was actually in the first paper to use the term "photometric redshift" (Koo, 1999). In the said paper titled "Near-Infrared Photometry of Distant Radio Galaxies: Spectral Flux Distributions and Redshift Estimates" (Puschell et al., 1982), the introduction of more studies of faint galaxies due to improvements in red and near-infrared sensitive detectors (Boynton, 1980; Gunn et al., 1981) meant a cheap and effective method for estimating redshifts of objects too faint for complete spectroscopy would have its advantages. In the spectral range mentioned above, models looking at the effects of stellar evolution on elliptical galaxies' observed colours suggested that their spectra' shape should remain constant for redshifts $z < 2$ (Bruzual A. et al., 1980). To test this modelled hypothesis using more faint data, Puschell devised the template fitting method in a photo-z frame.

At its most basic, the method seeks to best fit broadband data against known templates. These known templates would be a set of reference spectra where fluxes at a given filter are calculated and extrapolated across a given magnitude and redshift limit. The broadband data would then take the given redshift of the best fit template. This is accomplished through minimising the following χ^2 (Bolzonella et al., 2000):

$$\chi^2(z) = \sum_{i=1}^{N_{\text{filters}}} \left(\frac{F_{\text{obs},i} - bF_i(t,z)}{\sigma_i} \right)^2, \tag{3.18}$$

where σ_i are the uncertainty of the observed fluxes $F_{\text{obs},i}$ at given filter i wile $bF_i(t,z)$ is the template fluxes at a given reference flux b and redshift z and template t.

The ease of use of the template method has allowed for its application outside of just redshifts. One can also determine properties of celestial objects such as galaxy type by

making this a dependent variable in the templates used. Most importantly (and in contrast to empirical methods), the templates can be extrapolated to redshift and magnitude ranges outside the original template range but may lead to incorrect conclusions. With this said, the benefit of extrapolation has made template fitting the preferred method for new space surveys with a limited spectroscopic calibration set or the explorations of new regimes in existing surveys.

It is worth considering these templates in slightly more detail as they are essentially replacing the learning stage of the empirical methods, i.e. the templates are the model relating photometric data to redshift. The templates themselves come in 2 forms. Ether they are theoretical or empirical. Theoretical templates are created entirely from stellar population models. These models determine many galaxy features (such as stellar mass and element abundance). For understanding, we consider the most basic form of said models; the simple stellar population (SSP) (Renzini et al., 1981) how one gathers luminosity of different bands is outlined thoroughly by Maraston (2005). In this model, all stars are presumed to be of the same age t and with the same chemical composition (or metallicities) $[Y,Z]$ where Y,Z are the fraction of the total consisting of Helium or everything heavier (C, O, Mg, Si, Fe, etc., anything heavier than Helium is referred to as 'the metals') receptively. There is a 3rd component, X, which refers to the mass fraction of Hydrogen, but this can be implied as $X+Y+Z=1$. Sometimes a different metallicity notation is used, such as $[Z/H] = \log(Z/Z_{\odot}) - \log(H/H_{\odot}$ which refers to the total abundance of heavier elements concerning hydrogen relative to our sun. The total bolometric luminosity released by a said model star is split between main sequence stars (MS) and post-main sequence stars (PMS)

$$L^{\text{bol}}_{\text{SSP}}(t;[Y,Z]) = L^{\text{bol}}_{\text{MS}}(t;[Y,Z]) + L^{\text{bol}}_{\text{PMS}}(t;[Y,Z]). \tag{3.19}$$

The 2 luminosity contributions rely on completely different techniques to model. Main sequence stars span an extensive mass range, thus using isochrone synthesis (Chiosi et al., 1988). In this method, a mass-luminosity function $L(M,t)$ is adopted as well as an initial mass function $\Psi(M)$. The main sequence luminosity is then found by integrating over a series of mass bins along a curve on the Hertzsprung-Russell diagram, representing a population of stars of the same age but with different mass (an isochrone) (Kippenhahn et al., 2012)

$$L^{\text{bol}}_{\text{MS}}(t;[Y,Z]) = \int_{M_{\text{inf}}}^{M_{\text{TO}}} L(M,t)\Psi(M)dM, \tag{3.20}$$

where M_{inf} is the lower mass limit (usually $0.1M_{\odot}$) and $M_{TO}(t)$ the current turn off mass (the mass completing the hydrogen-burning phase). On the other hand, post-main sequence stars have very similar masses. In this case, the 'fuel consumption' approach is used (Maraston,

1998) is used. In this approach, the integration variable is no longer mass but 'fuel', defined as the amount of Hydrogen and/or Helium consumed by nuclear reactions in the post-main sequence phase. We use an evolutionary flux $b(t) = \Psi(M_{\rm TO})|dM_{\rm TO}/dt|$ which gives the rate at which stars evolve to any post main sequence phase j at age t and function of the amount of stellar mass that is to be converted into luminosity at each phase $\mathrm{Fuel}_j[M_{\rm TO}(t)]$

$$L^{\rm bol}_{\rm PMS}(t;[Y,Z]) = 9.75\times10^{10} b(t) \sum_j \mathrm{Fuel}_j[M_{\rm TO}(t)]. \tag{3.21}$$

The factor at the beginning of the equation is used as a unit conversion as $b(t)$ is given in years $\mathrm{Fuel}_j[M_{\rm TO}(t)]$ in solar masses and luminosity in solar units.

The choice of mass-luminosity relation and fuel consumption functions is paramount to producing luminosity and thus is 1 of the 3 'ingredients' that make the synthesis model (Maraston, 2005), 3 sets of matrices all containing chosen models that use variables from the previous ingredient. In this case, this ingredient is known as the *energetics*. The second ingredient beyond this thesis's scope is the *surface parameters*, which looks at the surface temperatures and gravities of the stars of different phases given their mass and luminosity as described in ingredient 1. Lastly, but most importantly for our use, is the third ingredient *transformation to observables* the spectra as a function of gravity and temperature (as produced from ingredient 2) to convert bolometric luminosity into usable SEDs. The transformation is completed either synthetically or empirically. For example, in the work of Maraston (2005) a synthetic library of individual star spectra is used (Lejeune et al., 1998) which is based on the modelling of star atmospheres. To have the correct spectra based on the ages, metallicities, surface temperatures and gravities generated through the primary model, a quadratic interpolation in $T_{\rm eff}, \log(g)$ and linear interpolation in $[Fe,H]$ takes place. This does not, however, cover the entire parameter space. For carbon-rich and oxygen-rich stars, an empirical library of observed spectra corrected for observation errors was used (Lançon and Wood, 2000). At this point, the luminosity's combined and intergalactic gas corrections are applied to give usable galaxy spectra such as that in Figure 3.10.

Alternatively, instead of creating the SEDs from models, one can use empirical data in the form of actually observed spectra of galaxies (Coleman et al., 1980). Although this removes the possibility of the synthetic modelling being inaccurate, new inaccuracies form through attempting to tie together observations at different wavelength ranges observed under different conditions. Other issues also become more prominent, the library of empirical observations does not cover all possible wavelengths, so modelling is required to extrapolate over entire wavelength ranges (Polletta et al., 2007). Furthermore, as these are accurate observations, one cannot control essential variables such as the galaxy's redshift of morphological type. With

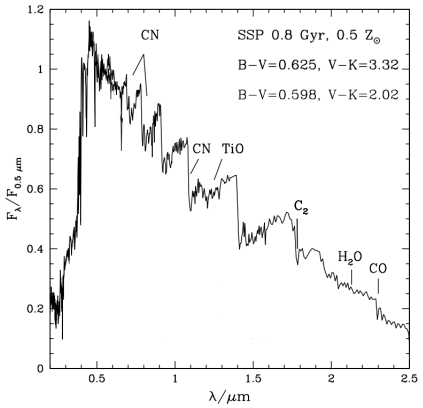


Figure 3.10: A synthetic SED of a galaxy containing a 0.8Gyr stellar population and metallicity of $Z = 0.5Z_{\odot}$ using a population synthesis model (Maraston, 2005).

that in mind, classifications of type and reversing redshifts to rest-frame are still required for optimum use. However, if template extraction is completed correctly, one will have a usable template set for interpreting galaxy spectra, such as the set in Figure 3.11.

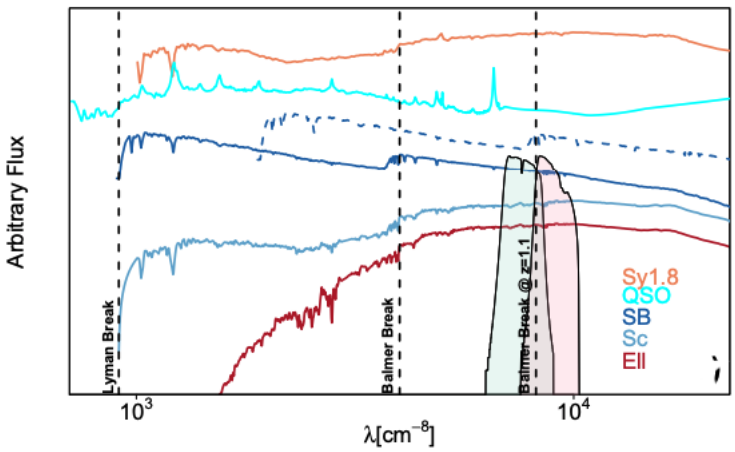
However, at this stage, the templates are not yet at the point of being model fluxes $F_i(t,z)$ as seen in equation 3.18. For the final conversion, we need to redshift the templates' luminosities $L_t[\nu]$ to every redshift within the desired range and integrate them over the band spectrum filters.

$$F_i(t,z) = \frac{1+z}{4\pi D_L^2(z)} \int_0^{\infty} \frac{d\nu}{\nu} L_t[\nu(1+z)] W_i(\nu) \tag{3.22}$$

Where $W_i(\nu)$ is the filter response as a function of frequency ν (Hogg et al., 2002). The filter response for filters i and z can. be seen in Figure 3.11. Equation 3.18 is only meant to represent the core principles of the template-based method and thus is in its most basic form. However, with the sheer quantity of different template-based codes, there are many variations in exactly how these templates are used. Take "EAZY: A Fast, Public Photometric Redshift Code" (Brammer et al., 2008) as an example. Instead of finding which individual template at a particular redshift best fits the data (which may lead to inaccuracies if the templates do not well represent the data), the code instead looks to find the best fitting set of coefficients from a linear combination of the templates.

$$T_z = \sum_i^N \alpha_i T_{z,i}, \tag{3.23}$$

where $T_{z,i}$ are individual templates at different redshifts and $\alpha_i \geq 0$ are the coefficients. N is the number of templates in the linear combination. This is usually 1,2, or all the templates in the list. Another differentiating factor is that the templates have error functions giving



centring

Figure 3.11: A typical template set consisting of SEDs from many types of galaxies (elliptical: Ell; Starburst: SB; spiral with small bulge: Sc and AGN ((luminous quasar: QSO; low luminosity obscured AGN: Sy1.8). Also included are typical features of SEDs, both at-rest frame and their place at $z = 1.1$ through dashed lines. Lastly, the transmission curves of the i, z filters are also included (Salvato et al., 2018).

certain wavelength regions different weights based on the accuracy of the template, which helps produce more realistic uncertainties. Alternatively, "The Zurich Extragalactic Bayesian Redshift Analyzer" (ZEBRA) (Feldmann et al., 2006) uses a machine-learning approach to generating templates. Using a training set of real SEDs of known spectroscopic redshift, a continuous space is created so that a set of basic templates are corrected at different redshifts to fit the space best (although obvious issues can apply with spectroscopic samples not representing the entire population). Other codes keep to basic template sets but tackle the dust attenuation issue (which can severely change the observed photometry). This can be achieved by using dust-corrected templates (Benitez, 2000) or by including dust as a free parameter (Bolzonella et al., 2000).

The fact that so many codes have taken dramatically different approaches to handling templates hints that there was/is a problem to be solved. Unfortunately, that is the case, and although template-fitting methods have their advantages, they also bring limitations that can lead to severe errors if not treated appropriately. The primary sources of error can be summarised into two main classes; colour/redshift degeneracies and template incompleteness (Benitez, 2000). The idea of degeneracy is that after minimising χ^2, there may be fluxes that fit more than one template. At this point, minor random errors in the reading of the observed flux would lead to objects being placed with the incorrect template and thus the incorrect redshift. This can be catastrophic if the possible templates are of wildly different redshifts (Sawicki et al., 1996).

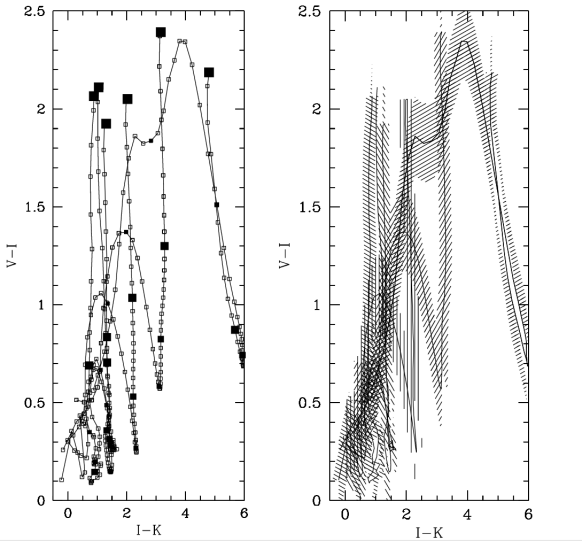


Figure 3.12: *left*: V-K vs I-K for a template sample in the $1 < z < 5$ redshift range. *right*: The same relationship with an applied 0.2 photometric error (Benitez, 2000).

The rate and severity in which degeneracies occur can be illustrated clearly through Figure 3.12 *left* which shows a colour-colour relationship (specifically V-I against I-K) for a sample of morphological types at redshifts $0 < z < 5$ the more significant the square, the larger the estimated redshift. When lines intersect, multiple estimations can be made from different templates. This is a degeneracy. The degeneracies can be significantly increased in their frequency through the implementation of error as seen in Figure 3.12 *right* where the lines have become 'blurred' at which point the probability of a catastrophic degeneracy is far more likely. As well as this blurring effect, increasing the number of lines or opportunities for crossing would lead to higher degeneracy. This would occur by increasing the number of lines and the redshift range. Such an error will not express itself randomly but may lead to a bias. It is expected that it would be far more likely that degeneracy leads to low redshift objects incorrectly identified as high redshift objects than in reverse, mainly because fewer high redshift objects are observed due to limiting magnitudes (this bias is looked into again when examining BPZ).

Unfortunately, the intuitive response of using more filters (theoretically, this would increase the amount of available information, thus showing differences that were not visible before) shows a slight improvement. For example, PCA analysis of the HDF-N photometric sample showed that no more than 3 parameters were needed to express the relevant information of seven filters (Connolly et al., 1995). That said, simple acts can counteract such degeneracy, such as limiting template size and redshift range to what is essential, thus limiting the opportunity for crossing, as mentioned above. Another possibility is to work out the best

magnitude limits to limit degeneracy. For example, this was achieved in photometric studies of the Hubble Deep Field (Sawicki et al., 1996), through Monte Carlo simulations, it was shown that catastrophic errors of redshifts are insignificant for bright objects but become significant for magnitudes around 27, so objects were only included with objects of magnitude 27 or brighter.

As mentioned above, template incompleteness is the other large area where error can occur. One can imagine a template set to be considered too 'sparse' if templates within it are too far apart in terms of type and redshift. If this is the case and the correct template is omitted, a redshift and spectral type may be chosen far away from the correct prediction simply because there were no templates closer to which to match to (Sawicki et al., 1996). The problem now moves away from the number of templates used to the accuracy of said templates. It has been shown that the template fitting photo-z method performs better when using empirical sets of templates than stellar population synthesis models (Walcher et al., 2010). The reason may seem simple. Instead of basing templates on observed spectra, synthetic models rely on the model having high accuracy. Unfortunately, this is not always the case. One example is UV spectra which are not well understood even for local galaxies (Sawicki et al., 1996), efforts have been made to include the effects of spectral evolution in their models, but these efforts do not include the effects of galactic self-absorption and other unknown effects.

To counter some of the main limitations of template-based methods, much focus has gone into the template allocation itself, especially on the problem of template incompleteness. We have already seen above some of the techniques used to fill the holes left by missing (or non-existing) templates, such as taking a linear combination of the templates instead of keeping them apart or 'correcting' their distributions based on SED training data. These are, in a sense interpolating the spaces between templates to produce more viable options. The level of interpolation can be pretty significant, for example, the LAPHARE template set (Ilbert et al., 2006). Similar to ZEBRA, a set of spectroscopic SEDs and redshifts are used, but unlike ZEBRA, they are not incorporated into a machine learning pipeline; instead, the features are used to interpolate a smaller set of templates linearly. In this case, an original set of 4 classically used templates (Coleman et al., 1980) (consisting of an E, Sbc, Scd and Im type galaxy derived from ultraviolet observations of nearby galaxies make up some of the components of base template sets still used today (Benitez, 2000)), as well as an extras starburst template (Kinney et al., 1996), were linearly interpolated using a set of galaxies from the VISMOS VLT Deep survey (Le Fèvre et al., 2005) to produce 62 new templates (over 12 times more than the original set). This same technique was later used to produce the commonly used COSMOS set as well (Ilbert et al., 2008). Other attempts to stretch the space of the templates have come through using free parameters and manipulating

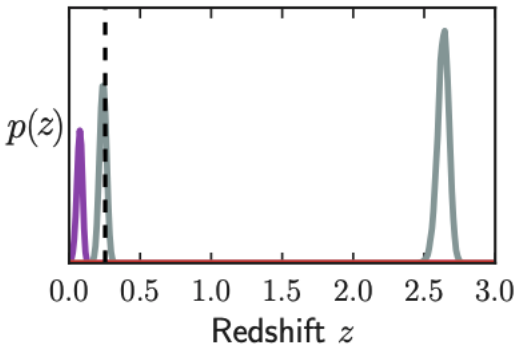


Figure 3.13: A typical p(z) generated from a simulated galaxy created from a template set, consisting of multiple peaks where a sub-peak agrees with the true redshift (dashed line) (Leistedt et al., 2016).

the template distributions (in a sense creating more templates), such as using dust as a free parameter, as seen above. As well as all these approaches work to mitigate one issue (template incompleteness), they, unfortunately, can also have the effect of exacerbating another, in this case, template degeneracy; By making the differences between templates finer and finer through the methods above you are in as sense adding more and more overlapping lines to the Figure 3.12 to such an extent that even with low errors in flux measurements one would still have multiple templates at multiple redshifts.

To deal with the much larger problem of template degeneracy, one may start by looking at changing the output of the entire process. Take, for instance, a situation that can occur as a direct consequence of degeneracy where if one was to treat χ^2 of the redshift template matching as the inverse to the probability (thus, the higher the $\chi^2(z)$, the lower the $p(z)$) of that redshift being the redshift of the photometric source then due to degeneracy one could have multiple peaks of high probability where the highest peak may not even be the accurate redshift. An example can be seen in Figure 3.13. Under the outputs we have seen up to this point, one would receive a single photo-z estimation based on the redshift value corresponding to the highest probability (lowest χ^2), which in this case would be catastrophically wrong. Thus a new output is required that somehow incorporates information from the sub-peaks. The first attempt at such incorporation was from the work of SubbaRao et al. (1996) who used the mean redshift of the whole range of possible redshifts instead of the maximum probability. This allows for the variance to also be outputted, giving an idea of the uncertainty of the result. If the said approach were applied to Figure 3.13, one would get a mean Redshift much lower than the maximum estimate with a large enough uncertainty to include the accurate redshift in its error. Such a technique's success led to the expected outcome of most redshifts codes today, starting with the work of Arnouts et al. (1999) the idea being that one does

produce a single estimate but instead produces the probability density function (PDF) $p(z)$ such as the one seen in Figure 3.13. In this sense, no information is lost, which becomes even more critical when considering summary statistics of complete surveys, as we will see below. The mean Redshift technique was closer to this idea than first thought, effectively restraining the PDF to a Gaussian of width and peak dictated by the mean and variance. Using the entire PDF photo-z estimates of quasars from the SDSS optic sample, one could remove all catastrophic errors (Lima et al., 2008). Of course, there are still situations where a single number estimation is required. This case needs new ways of extracting information from the $p(z)$. One such way is taking random samples of the $p(z)$ and choosing the peak redshift based on these samples (Wittman, 2009). This approach showed much less scatter in z_p vs $z_{\rm spec}$ diagrams, which would significantly reduce systematic for weak lensing studies. Other techniques that use the whole distribution but will not be covered in detail include post-selecting peak techniques (Lima et al., 2008) and risk parameter techniques (Tanaka et al., 2018).

Although changing focus to full PDFs has apparent benefits, there are a couple more subtle issues that may seem detached from the primary process. The first is storage. Depending on the size of the Redshift bins used to construct the pdf, one can quickly require anything from 50 to 1000 times more data points per galaxy which can become detrimental for multi-million Galaxy surveys. Attempts to solve the issue have been both mathematical and technical, either through representing the data in other ways such as Gaussian Basis functions (Kind and Brunner, 2014) or through the inclusion of more storage of the data reduction techniques (different file formats etc.). The other issue is validation. It is elementary to validate a photo-z technique if one is comparing a single z_p value to a single $z_{\rm spec}$ value but much more difficult if one has a whole distribution. One such attempt is the continuously ranked probability score (Polsterer et al., 2016) which measures how well the $z_{\rm spec}$ is captured within the curve of the $p(z)$

3.3 Summary

The above chapter looked at one of the core aspects of weak lensing surveys, and one covered for the remainder of this thesis: photometric redshifts. The weak lensing survey considered distortions of sources due to matter between observer and source. A distortion that would be meaningless without an idea of the distance or redshift between them. The redshift of the sources is so paramount to weak lensing surveys that the sensitivity of the cosmic power spectrum to the mean redshift of the survey is as sensitive as any of the cosmological parameters one is trying to determine (Jain and Seljak, 1997). In fact, the mean redshift

is so detrimental that a rough estimate of the convergence power spectrum can be roughly estimated to be more than proportional to the mean $\bar{z}\, P_\kappa(\ell \sim 1000) \propto \Delta_{de}^{-3.5} \sigma_8^{2.9} 1\bar{.}6|\omega|^{0.31}$. One of the accepted best avenues for finding accurate redshifts is by examining observed galaxy fluxes (or magnitudes) from sources, as these can be directly linked to redshifts due to district spectral features. The most accurate method is through spectroscopic redshifts in which one observes very fine wavelength bins leading to full spectral energy distribution. Although accurate, the method is costly and limited to brighter sources, so not suitable for the millions of faint sources typical of a weak lensing survey.

Instead, one may use photometric redshifts, a technique similar to the above but with much broader wavelength filters. Due to their broadness, one cannot rely on individual fine spectral features like the spectroscopic case. Instead, one must use broader features or otherwise unknown links between flux's/magitudes (or their ratios/differences in the forms of colours) and redshift, which is achieved through data-driven empirical techniques (both supervised and unsupervised machine learning). Although empirical techniques are accurate compared to template methods, they do not work well in the conditions in many weak lensing surveys, which is a broad range of possible redshifts in areas with very little spectroscopic training data. Instead, one may use template-based methods, which take advantage of the fact that the galaxies of different mythologies contain different spectral features. By having templates that match these different morphologies, one can mitigate some of the inaccuracy of the input data. These templates are formed from either synthetic models (that may have inaccuracies in their modelling) or empirically gathered from observation (which may not sufficiently represent the input survey). Although more accurate (than data-driven methods) in spectroscopically poor regions (and thus valuable for weak lensing surveys), they have their weaknesses in the form of template incompleteness and degeneracies where the former can be reduced through different template interpolation techniques as seen in the many different modern photo-z codes. Although changing from single photo-z estimates to complete probability distributions does work to address some of the template degeneracy concerns, to tackle the issue directly, one must take a slight detour to ask 'what does it mean for a certain redshift to be most probable?'.

Chapter 4

Bayesian Redshift Statistics

4.1 Statistical Theory

4.1.1 Frequentist Inference

To understand a new way of statistical thinking, we must shore up our grasp on the current way. We have mentioned the idea of the χ^2 being inversely proportional to the probability. This is only true in the vagueness sense of which one rises and the other falls. The χ^2 we witnessed in equation 3.18 is a simplified case of a more general equation. In this simplified version, all observations are independent and thus have errors with no covariance. If we were to lift these restrictions, we would return to the whole generalised equation as

$$\chi^2 = (\hat{D} - D(\theta))^T \Sigma (\hat{D} - D(\theta)), \tag{4.1}$$

where $\hat{D} = \{y_1, ..., y_N\}$ is the data vector, $D(\theta) = y_1(\theta), ..., y_N(\theta)$ the model vector, θ is the changing parameter and Σ the covariance . Of course, in our case, the data vector would be the filter fluxes, the model being the expected template fluxes at different redshifts and the redshift parameter. To return to equation 3.18 one simply removes the covariance by setting $\Sigma = \text{diag}(\sigma_1, ..., \sigma_N)$ which simply becomes a vector of the errors on the filter fluxes. To link this value to a probability, one must make assumptions about the form of the data we are collecting. The main assumption is that the value of the filter flux y_i collected can be approximated to a Gaussian distribution around the mean value of $\bar{y}_i$ and variance σ_i. This assumption usually holds for astronomical data. Due to the large amount of data being collected, gaussianity is implied by the central limit theorem (Shariff, 2017). This means one not only has a value of how close this is to the model but an idea of how likely it was that

these input values represent the true values of the source

$$p(\hat{D}|\theta) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\chi^2\right). \tag{4.2}$$

This probability has the particular name 'likelihood', and its true meaning can be taken from its symbolic form $p(\hat{D}|\theta)$ as is to say is the probability of observing the data $\hat{D}$ *given* the parameter θ is true. This gives us a much more intuitive idea of what this probability form is trying to achieve.

Consider these 2 scenarios, starting with a low error faint observation of a source. If we assume (or 'given') that we know the source is at a low redshift where our templates (models) tell us a source will give high fluxes for this redshift, there is a low probability of reproducing that result as the lack of error in the value means it would be very rare for the data to vary that far away from the actual high flux mean. Whereas a second example, a source is producing fluxes that completely agree with the model under the assumption that it has a particular redshift but has an extremely high error associated with that. The likelihood here is also low (in fact, the likelihood would be close to uniform regardless of flux value) mainly because here it would take many multiples of new observations to receive that accurate flux again due to the highly varying nature of the significant error.

In both cases, we were using the language of 'reproducing' the result or 'taking multiple observations' because if we consider the likelihood to be our main pointer of probability, then we are taking what is known as 'frequentist' statistics.

Statistics of this kind get their name because it says the probability of something occurring is referencing the *ratio* (of frequency) of that event occurring compared to not occurring if the experiment was repeated an infinite amount of times. For example, if one was to say that a coin flip has a 50% likelihood of being heads, what one is saying from the frequentist point of view is that if we were to flip a coin an infinite amount of times repeatedly, half of the times, the coin will be heads. The idea of a frequentist approach being scientifically used came into prominence with the rise of High Energy Physics and particle collision, where 1000s of replicated collisions are taking place every second. Under these circumstances, one can refer to the probability of a particular particle being created every 100 collisions in a complete frequentist framework, and due to the ease with which likelihoods can be produced, this was and still is the general approach in these realms. Our typical intuitive idea of probability seems frequentist at first, much like the coin example above, but this is not the way we usually think or ideally want to think and certainly does not apply to situations in which an experiment cannot be easily repeated. One such example is that of cosmology. We only have one universe and thus one data source to observe. Under these circumstances,

what does 'frequency' or repeated experiments even mean? Under these circumstances, a different method of statistics is required.

4.1.2 Bayesian Inference

The 'new' statistics, which we will explore and later apply to our photometric redshift problems, are a version of statistics that precedes the frequentist approach. Derived initially in the 1740s by Reverend Thomas Bayes and later published after his death by Richard Price (Bellhouse, 2004). The Bayesian approach takes probability away from this idea of repeating experiments and frequencies of occurring to something much more in tune with how the scientific method works. Instead, it moves it to a statement of 'belief' (D'Agostini, 2005). At first, this may seem contradictory, but it is more in tune with how we interpret probability in the real world. Let us take the example of a flipping coin. When we state the probability of getting heads on the next coin flip is 50% instead of saying this means that if we were to flip it an infinite amount of times, it would be heads half as much of the time (less intuitive as we only have the single flip). Instead, we are saying that based on the 1000s of coin flips we have seen, we believe the next event of heads is just as likely as tails. The significant idea here is that we use prior knowledge to shape our probability and use incoming data to change that probability if needed. For example, if the next 50 flips were to head, our 'belief' that the next flip is a 50:50 split (or, more precisely, that we have a fair coin) may begin to drop. These are all scenarios allowed through Bayesian statistics.

Of course, this is not just a philosophical reinterpretation of probability for one to keep in mind when constructing experiments and their statistical analysis. A mathematical framework allows it to differ from the frequentist method. The mathematical formalism was first independently developed by Pierre-Simon Laplace in 1774 (Zeger, 2012) and is based on considering events (A and B). A common misconception of probability theory is that one can simply write the probability of both events occurring as $p(A \cap B) = p(A)p(B)$ where $p(A \cap B), p(A)$ and $p(B)$ the joint probability of A *and* B occurring, the probability of A occurring and the probability of B occurring respectively. This is only a special case of the real, more general equation $p(A,B) = p(A|B)p(B)$ where $p(A|B)$ is the probability of A occurring *given* B has occurred. For simplicity, we have used (and will continue to use) an alternative formalism for joint probability of $p(A \cap B) = p(A,B)$ mainly due to the rise of events one can have in a single probability going forward. The conception (and thus the special case) comes from the fact that when one first learns probability theory, they usually use a situation in which 2 events are completely independent of each other, and in that situation, the first equation is correct as $p(A|B) = p(A)$. The second piece to revealing the Bayesian framework is by using the transitivity of joint probability, which means the

probability of A and B happening is the same as the probability of B and A. This may seem linguistically trivial, but it opens the door to the rest of the derivation,

$$p(A,B) = p(B,A), \tag{4.3}$$

$$p(A|B)p(B) = p(B|A)p(A), \tag{4.4}$$

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}. \tag{4.5}$$

Equation 4.5 is the famously known 'Bayes theorem', giving us a new interpretation of statistics. It may not be self-evident *how* that is the case, but it will become more evident as we replace A and B with more contextual events. For example, consider B to be $\hat{D}$ the data just observed from an experiment and A to be θ a particular parameter or hypothesis connected to one is observing. Under these transformations, we now have the following:

$$p(\theta|\hat{D}) = \frac{p(\hat{D}|\theta)p(\theta)}{p(\hat{D})}, \tag{4.6}$$

this is the more useful equation, especially in the field of parameter inference. To understand its significance, we shall break it down into individual terms.

- $p(\hat{D}|\theta)$ - the *likelihood*, the probability that the data will be generated given the hypothesis (or parameter) is true, the same term maximised in frequentist statistical analysis
- $p(\hat{D})$ - the *evidence*, the probability of the data unconnected to the theory
- $p(\theta)$ - the *prior*, the probability of the parameters based on all knowledge before the experiment has taken place and data observed
- $p(\theta|\hat{D})$ - the *posterior*, the probability of the parameter or hypothesis given the data observed

It is the last 2 terms that show the divergence between Bayesian and frequentist approaches. The likelihood is no longer the goal quantity but a component of the overall goal, which is the posterior, i.e. my 'belief' in this hypothesis given the new data I have just observed. These beliefs are not in a vacuum but are influenced by prior beliefs. Thus, in the case of the coin, flipping 1 head would not affect our posterior significantly due to prior knowledge of the coin's fairness up to that point.

The subtle difference between the likelihood and the posterior and the evidence's meaning can feel very abstract. With that in mind, to illustrate not only there mathematical meanings

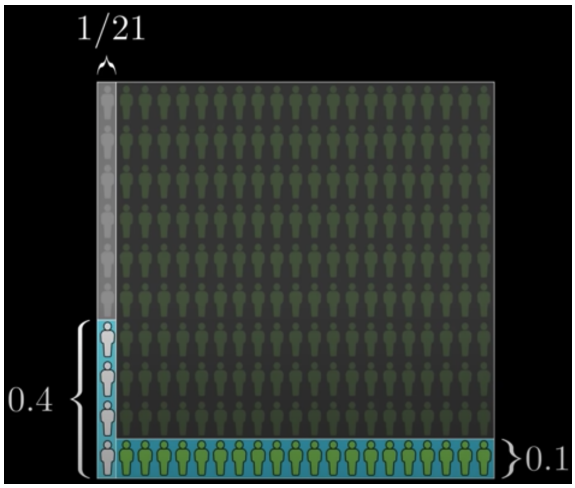


Figure 4.1: An illustration of a sample of 210 people, 10 of which are librarians (grey) and 200 of which are farmers (green), those in blue are those that match the given description of the man (Sanderson, 2019)

of these terms and how this can help change our way of thinking of probability can look to the behavioural psychology study of Daniel Kahneman and Amos Tversky and Kahneman (1974) which was later illustrated for widespread consumption (Sanderson, 2019). The experiment was a series of people who were told the following (paraphrased) 'A man was randomly chosen from the American population. Using the description below, do you believe this man is a librarian or a farmer?'. They were then described the man with words one may stereotypically associate with a librarian, not a farmer. People overwhelmingly chose a librarian that Daniel and Amos deemed irrational when determining. The reason is that people did not consider the prior knowledge that the ratio of farmers to librarians in America is 20:1 to librarians.

Figure 4.1 offers clarity to the distinction. Illustrated here is a population of 210 people, of which 200 are farmers (green), and 10 are librarians (blue). The highlighted people in blue represent those whom the description would match. People determining that the person is a librarian, not a farmer did not by comparing their likelihoods i.e. $p(\text{description}|\text{librarian}) = 0.4$ the probability that given the person *is a librarian* the description is more likely to description than if the person was a farmer $p(\text{description}|\text{farmer}) = 0.1$. Instead, we want the probability of the person being a librarian given the description, i.e., the posterior

$$p(\text{librarian}|\text{description}) = \frac{p(\text{description}|\text{librarian})p(\text{librarian})}{p(\text{description})}. \tag{4.7}$$

We must now incorporate the prior which in this case is $p(\text{librarian}) = 1/21$ (the 20:1) ratio (or the proportion of grey to green people) while the evidence $p(\text{description})$ is the num-

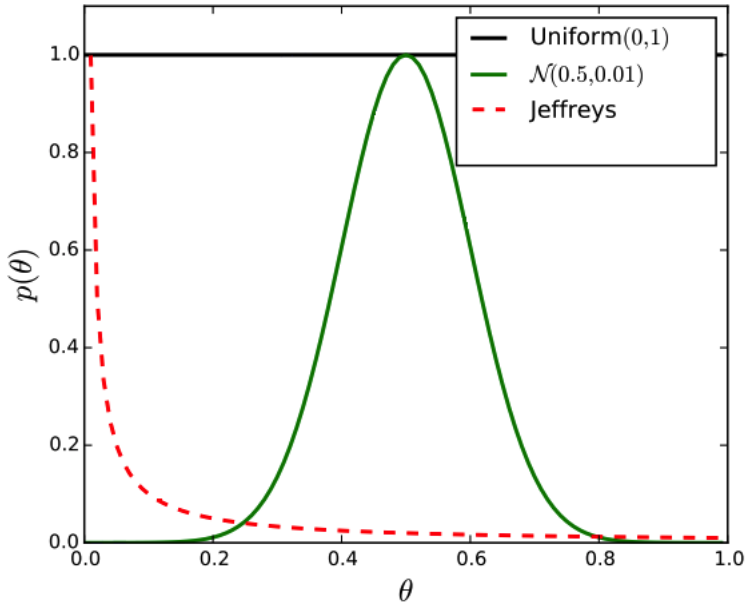


Figure 4.2: Examples of gaussian, uniform and Jeffreys priors (Shariff, 2017).

ber of people to which the description matches them i.e., everyone highlighted in blue $p(\text{description}) = p(\text{librarian})p(\text{description}|\text{librarian}) + p(\text{farmer})p(\text{description}|\text{farmer}) = 24/210$ giving us a true posterior of 16% or $4/24$ which makes sense as in our population of 210 people, the description matched 24 of them to which 4 were librarians. As you can see, even with the likelihood being 4 times higher, our priors can hugely affect the outcome. Which they should, our beliefs are not founded on the data in front of us but on all the data of previous experiments.

4.1.3 Choice Of Priors

This idea of constantly updating beliefs so that the current posterior becomes the following experiments prior means likelihoods can be constantly multiplied onto each other. As long as the prior is non-zero in regions where the likelihood is significant, repetitions of the same experiment will lead to the posterior converging to a single function (Leclercq, 2015), much as our belief in a hypothesis would also reach a convergence point. With that in mind choosing the correct prior is of high importance, which usually comes in 2 forms, either informative or uninformative.

An informative prior would be one based on prior knowledge that has been collected and can be the posterior of a recent experiment or be synthetic based on outside knowledge, such as boundary conditions to the accepted answer, or as we saw above, based on population numbers. A classic example would be to go back to the coin-flipping experiment one may

use a Gaussian prior of $p(\text{flipping heads} \sim \mathcal{N}(0.5, 0.1^2))$ (Shariff, 2017) where you are very sure that the coin is fair (as shown through the low variance around 0.5). These priors are examples in Figure 4.2. Due to the very nature of being based on the knowledge of that experiment, informative priors are usually highly context-specific, which is much more general and exposes some of the nuances of prior choice that come through non-informative priors. These priors are supposed to state that we have no prior information on the correct parameter values. The most obvious would be a uniform prior $p(\theta) = U[\theta_{\min}, \theta_{\max}]$,

$$U[\theta_{\min}, \theta_{\max}] = \begin{cases} \frac{1}{\theta_{\max} - \theta_{\min}}, & \theta_{\min} \leq \theta \leq \theta_{\max} \\ 0, & \text{otherwise.} \end{cases} \tag{4.8}$$

This is to say that within the boundary conditions, every θ value is just as likely as any other. A perfect example of its use is back to our coin, where if we wanted to not presume at all that it was fair or not, we would attribute the prior probability of getting heads to $p(\text{flipping heads}) = U[0,1]$. At first glance, this should be the end of the discussion as what can be less informative than a prior that treats all possible θ values as equally possible? Unfortunately, issues begin to arise if one needs to re-parameterise. Consider the following thought experiment by Alsing (2017); You are trying to find the value of some parameter θ, as you have no prior knowledge of θ other than it is a probability you chose to assign a uniform prior of $p(\theta) = U[0,1]$ (just as we did previously with the coin). Later on, it is decide to re-parameterised to a new parameter of interest $\lambda = \ln(\theta/(1-\theta))$ as you equally do not know λ it seems sensible to use the same uniform prior of $p(\lambda) = U[0,1]$. However, due to the relationship between λ and θ you have unintentionally changed the prior of θ to $p(\theta) \propto (1-2\theta)/(\theta(1-\theta))$ as $p(\theta) = p(\lambda) \times |d\lambda/d\theta|$. The more parameterised, complicated and interconnected a scientific problem becomes, the harder it is to blindly attribute parameters to uniform priors without potentially biasing other aspects of the analysis. Thus different methods were created to have priors potentially independent of the parameterisation. One such example is the Jeffreys prior (Jeffreys, 1998) which formally is defined as

$$p(\theta) \propto \sqrt{\mathscr{F}(\theta)}, \tag{4.9}$$

where $F(\theta)$ is the Fisher information, a measure of how much information one can get on the parameter θ given the data available

$$\mathscr{F}(\theta) = -\int p(\hat{D}|\theta) \frac{\partial^2 \log p(\hat{D}|\theta)}{\partial \theta^2} d\theta. \tag{4.10}$$

The exact details of the Fisher information are beyond this thesis but at its most basic the Jeffreys prior is uniform in log space i.e. $p(\log(\theta)) = \text{constant}$ and $p(\theta) \propto 1/\theta$. This can be especially important if one does not know the scale on which θ exists (i.e. very small $10 * -10$ or very big 10^10). If one used a very wide uniform prior, then most of the probability will lie with the larger numbers, i.e. there is more probability for $\theta > 1$ than $0 \leq \theta \leq 1$). Whereas the Jeffreys prior would be the same weight at all magnitudes. Jeffreys is a 1-dimensional version of a more general reference prior; however, finding such priors for all problems can be taxing recurring the solving of calculus-of-variation problems and thus is not covered in more detail. Alternatively, there are maximum entropy priors seen as computationally less intense. They state that the least informative prior would maximise entropy given a set of prior constraints where the entropy of the probability distribution is given as

$$H(p) = -\int p(x) \log p(x) dx. \tag{4.11}$$

This allowed for a constraint-focused instead of a model-focused approach and removed the options of choice to something more principled in what is least informatively prior. If the only constraint is $\int p(x)dx = 1$ (a constraint required for this method to work), the maximum entropy prior is uniform. The maximum entropy prior is a gaussian if the mean and variance are fixed. However, unlike above, they are not insensitive to parameterisation, so they should be used with care.

4.1.4 Hierarchical Models and Marginalisation

The above showed how Bayesian statistics changes the way we consider probability and how best to incorporate prior knowledge in a way not suitable to the frequentist approach. That said, it is always possible to include prior knowledge within the frequentist analysis. For example, this could be achieved by adding a weighting function (say one that weighted towards the outcomes of previous experiments) *within* your likelihood function. This mathematically is true and would lead to a likelihood result proportional to its posterior counterpart. Unfortunately, it is principally dubious; by modifying your likelihood, you, in a sense, modify your model and thus the theory behind it, implying that experiments (and the random variables they produce) are not giving independent results but are dependent on the experiments that have come before it.

Bayesian thinking allows for a clear answer to this issue. The underlying model remains intact with the modelled likelihood and the prior knowledge as separate functions that only

join to give a posterior. This same trail of thinking can be applied to another topic in which both approaches differ; the dealing and propagating of uncertainties.

In its most basic form were are constantly dealing with a set of observed data $\hat{D}$ and a model to explain that data with particular parameters θ. Under the frequentist framework, one would maximise the likelihood $p(\hat{D}|\theta)$ to find which parameters best produce the observed data. Other than the misgiving mentioned above, if this is the function we want to maximise, another question or issue arises. What does the rest of the likelihood distribution represent? Why would it not produce the same data every time? Why is there a probability attached at all?

Let us remove the idea of the model having some underlying uncertainty by making it based exclusively on classical physics. Under this assumption, the only form of uncertainty and thus the need for distribution comes from how we observe that data, for example, the quality of our instrumentation leading to error. The frequentist approach includes measurement error within the likelihood function, which for small enough models is fine (an utterly working example being the photometric redshift likelihood function used for both Bayesian and frequentist approaches alike). However, this can lead to 2 issues, the first being the principled issue mentioned above; should the model depend on my instrumentation? If I upgraded my instrumentation, this would potentially sharpen the likelihood. The physics model now seems to be affected by my choices. On the other hand, the mathematical issue deals with what happens when those uncertainties affect different parameters (say, one within a sin function, for example). Such propagation requires mathematical rigour and attention to ensure they affect the output data correctly.

The Bayesian answer, as with prior knowledge, is to keep the observed data's uncertainties as a separate entity. However, seeing as the terms of Bayes Theorem have already been accounted for, how does one do this? One uses a hierarchical model. Unlike above, where one goes from parameters to observed data through one likelihood function, now we have multiple likelihood functions. The first is the physical model $p(D|\theta)$ linking the *true* data D to the parameters and the second is a completely separate instrumentation model $p(\hat{D}|D)$ linking the true data D to the observed data $\hat{D}$. We can rewrite the Bayes theorem so that to 2 things we now care to find are parameters and the correct data given the observed data

$$p(D,\theta|\hat{D}) \propto p(\hat{D}|D,\theta)p(D,\theta). \tag{4.12}$$

Proportionality refers to the fact that the evidence has been ignored. This is allowed as direct comparisons of 2 posteriors cancel out the denominator. It is, in a sense, a normalisation constant. We can use the multiplicative law to rewrite the prior in terms of the physical model, i.e. $p(D,\theta) = p(D|\theta)p(\theta)$. Through the fact that we have already stated that there is

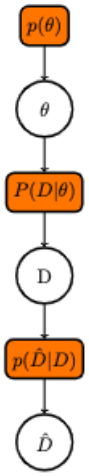


Figure 4.3: An example of a basic Bayesian hierarchical model based on the toy model used above, consisting of 3 main components: the physical model likelihood; the instrumentation error likelihood; and the prior.

no direct model linking observed data to parameters (only observed data to real data), we can also rewrite the likelihood in terms of the instrumentation model $p(\hat{D}|D,\theta) = p(\hat{D}|D)$. Leading to our joint posterior

$$p(D,\theta|\hat{D}) \propto p(\hat{D}|D)p(D|\theta)p(\theta), \tag{4.13}$$

as models become enlarged with more and more variables. Hierarchical models are best described through directed acyclic graphs as seen in Figure 4.3 one has a path from prior knowledge to observed data for which all models can be accounted for.

Of course, there is one remaining issue, we don't want the joint posterior $p(\theta,D|\hat{D})$ we want the posterior to the parameters alone $p(\theta|\hat{D})$ and they are not equivalent. Under these circumstances, the actual data D becomes known as a nuisance parameter, something we want to take account of but not includes in the final distribution (i.e. we ultimately do not care what the actual data should have been, that is just a comment on our instruments, not the physical model). This becomes especially important for much larger Bayesian hierarchical models with many different uncertainties and thus many nuisance parameters. To deal with them, we use a method known as marginalisation to integrate all possible values of that variable weighted by its probability of occurring

$$p(\theta|\hat{D}) \propto \int dD p(\theta,D|\hat{D}). \tag{4.14}$$

Under this framework, we now have a consistent approach that keeps models independent, and uncertainties accounted for. Although, for a complete picture, one may still wonder why there is a model likelihood if the parameters would produce the same data every time (under our classical view). Under these circumstances, our model likelihood would become a direct delta function and thus is still viable for calculation.

4.2 Application

4.2.1 Computation

So with the apparent benefits and philosophical consistencies, Bayesian statistics offers advantages that should not have been ignored. However, they were with frequentist statistics still being the overall analysis used in modern science. For starters, as the number of observations tends to infinity, the two methods converge in their conclusions negating the said benefits. One concern for Bayesians is in this choice of prior, which can affect your overall result, an effect frequentists can ignore outright. However, this argument can be used in reverse, with frequentists having to take great care in propagating their errors correctly, especially with implicit Gaussianity of χ^2, something Bayesians do not have to worry about. Due to their hierarchical modelling, we can propagate the errors with whatever distribution we require. There is no constraint forcing Gaussianity. The area in which Bayesian statistics falls short and thus is ignored in favour of its frequentist counterpart is computation.

If we were to take the basic Bayesian hierarchical model of Figure 4.3 if one were to compute the frequentist Gaussian likelihood, one would have an analytical function to simply plot. With the Bayesian approach, we must multiply the individual components (prior, physical model, instrumentation error) and integrate all possible data combinations to evaluate the posterior. This can cause 2 major computational problems. The first is that with that many components, there is no guarantee of an easy analytical function to compute. Efforts can be made to simplify some mathematics by using conjugate priors. These are the priors of which their distribution was chosen so that they can easily be multiplied by the likelihood. One such example would be the conjugate before a Gaussian likelihood would be a Gaussian prior, as when their product (and thus proportional to the posterior) would also be Gaussian. The issue with this approach is that you now choose your prior for convenience as opposed to any principled reason, and even then, not all easy computations can be guaranteed. Some posterior functions are not analytic at all and require simulation to compute. Each one of these computations can be costly.

Figure 4.4: An illustration of samples being taken from a normalised distribution proportional to the distribution itself regardless of normalisation, these samples can then be used for statistical analysis (Rocca, 2019).

The second computation issue comes from what we are trying to extract, saying we had a posterior (and thus a likelihood) that was reasonably easy to compute individually. The frequentist approach only requires the likelihood to be maximised. Thus many different optimisation tools can be used. Bayesian, however, desires the entire distribution to be evaluated. One can imagine simply splitting the distribution into N equally spaced points on the parameters axis and evaluating each point to plot the distribution. This does work and is entirely viable for 1-dimensional problems. However, as seen with the hierarchical model, nuisance parameters add up. In the case of complete cosmological analysis, there are multiple cosmological parameters of which one would be evaluated. If there are k parameters, that would be at least N^k evaluations. Computational time can quickly skyrocket to untenable in what is known as the 'curse of dimensionality (Bellman, 1958).

Due to this curse, one cannot simply evaluate every point. Even at low dimensions, the issue resides in finding the balance between a possible computational number of points and the required level of precision. Instead, one needs to employ a sampling method in which samples of the distributions are taken (where each sample would consist of the parameter values), and the number of samples taken within any $\theta_{\text{max}} - \theta_{\text{min}}$ region is proportional to the posterior distributions within that region. Put regions of the high posterior will have had more samples obtained than regions of the low posterior, as shown in Figure 4.4. Besides being computationally less expensive than a complete distribution evaluation, the sampling method also contains other valuable by-products. Say one has a table of samples consisting of columns of parameter values where every row is a sample. The first thing to notice is that, as shown in Figure 4.4, there is no retaining posterior information. The number of samples only implies this. Thus there is no need to normalisation the distribution as this is done automatically. Secondly, under these circumstances, statistical analysis becomes a much easier task. If you wish to marginalise a particular parameter, one no longer needs to integrate over the entire distribution. They need to ignore all columns *other* than their

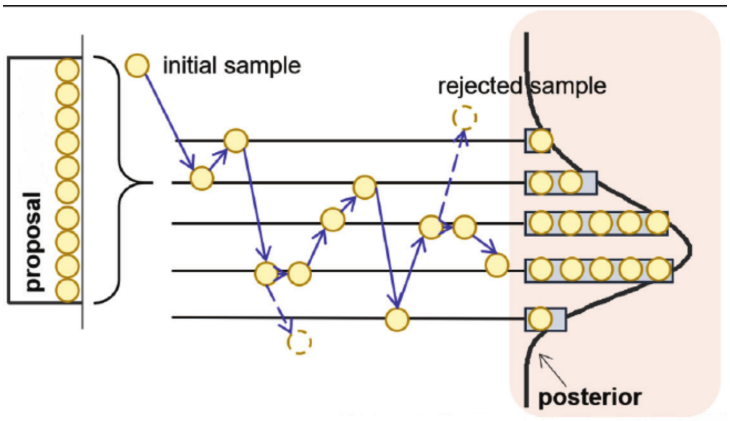


Figure 4.5: An example of the MCMC MH algorithm, containing a proposal distribution and the uptake of samples both rejected and accepted to the proportions of the posterior distribution (Dong et al., 2020).

parameter of interest. With sampling outputs, the checks of quantities such as mean, variance and correlation with other parameters become a trivial exercise of data manipulation instead of computation.

Thankfully such computational methods exist that allow us to sample the distribution. The most efficient family of known methods are Markov Chain Monte Carlo (MCMC) methods in which one collects a 'chain' of samples of which the following chain sample in the chain θ_{i+1} is probabilistically linked to the current sample in the chain θ_i. The simplest and most popular version of this method is The Metropolis-Hastings (MH) algorithm (Hastings, 1970)(Metropolis et al., 1953). The algorithm can be seen illustrated in Figure 4.5 and contains the following steps.

1. Randomly select an initial sample θ_i from a distribution (such as the prior $p(\theta)$)
2. choose a proposal sample θ_{i+1} based on a proposal distribution depended on the previous sample $p(\theta_{i+1}|\theta_i)$
3. selected a random number r from a uniform distribution of $U[0,1]$
4. if $r < Q(\theta_i, \theta_{i+1})$ where $Q(\theta_i, \theta_i + 1) = p(\theta_{i+1}|\hat{D})/p(\theta_i|\hat{D})$ is the acceptance criteria, accept the sample
5. if $r > Q(\theta_i, \theta_{i+1})$ reject the sample and declare the new sample $\theta_{i+1} = \theta_i$
6. repeat steps 2-6 for the next sample θ_{i+2} up to θ_{N_s} where N_s is number of samples.

As you can see, the acceptance criteria is a ratio of the 2 posteriors. This ensures that the algorithm does not just become an optimisation procedure (if that was the case, only higher

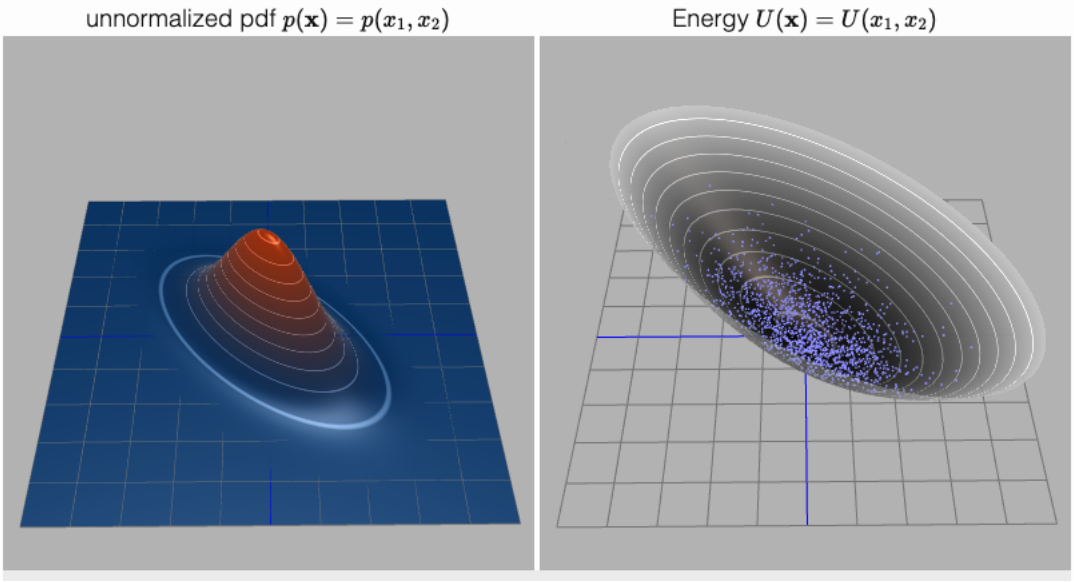


Figure 4.6: An example of a simple 2-dimensional posterior distribution on the left being converted to a potential energy distribution on the right of which samples are drawn (pink) (Rogozhnikov, 2016).

and higher posteriors would be chosen) but always has a chance of sampling a low posterior region if a tiny chance. The method's weaknesses come through mainly in its proposal distribution $p(\theta_{i+1}|\theta_i)$ this is where the ad hoc choices are made to ensure convergence of the whole distribution. It's usual to see a Gaussian distribution be used in these circumstances such as $p(\theta_{i+1}|\theta_i) = \mathcal{N}(\theta_i, \Sigma)$ one can that the size of the jump from one sample to the next is dependent on Σ. It is this choice that requires care. Too small a jump, the procedure will take too long to cover their entire distribution. Too large, and the differences in the posterior between samples would be so large that one will face a large number of rejections and be incredibly inefficient. This acceptance rate becomes a prominent issue as we move to higher dimensions to which even small jumps can lead to large changes in the posterior. This MH becomes impractical at dimensions of ≥ 10. For these situations, alternative MCMCs are required, such as Hamiltonian Monte Carlo, which treats samples slightly like travelling particles where their potential energy is determined by the probability distribution and the kinetic energy determined by a chosen 'temperature', as can be seen in Figure 4.6. This method allows us to look at much larger dimensional problems (with dimensions $> 10^6$)) (Neal, 2012) but with a cost, computing individual jumps is much more difficult, requiring the gradients of the posterior, a quantity that cannot always be assured.

An alternative approach (and the sampling approach used for the majority of this thesis) for jumping from sample to sample does not require the full posterior. Instead, one can use conditional probability (in some cases, such as in 5 the conditionals are exponentially easier

to compute). To explain the algorithm, we will consider a simple probability distribution $p(x_1, x_2)$. This may seem like an oversimplification because it is not conditioned on data nor contains a higher amount of dimensions; however, the algorithm below would apply the same for both cases. The algorithm is illustrated in Figure 4.7

1. choose a random initial sample $x^{(t)} = (x_1^{(t)}, x_2^{(t)})$
2. take a new sample along the x_1 axis by random choosing from a point along the conditional distribution $x_1^{(t+1)} \sim p(x_1|x_2^{(t)})$ so that the new sample is $x^{(t+1)} = (x_1^{(t+1)}, x_2^{(t)})$
3. take a another sample along the opposing x_2 axis so that the next sample is $x^{(t+2)} = (x_1^{(t+1)}, x_2^{(t+2)} \sim p(x_2|x_1^{(t+1)}))$
4. repeat stages 2 and 3 until we have N_s samples

The only alterations to include the points mentioned earlier are that one would include the observed data and loop through all possible dimensions so that the conditional probability used is $p(x_i|x_{j\neq i}, \hat{D})$. Since the conditional probabilities are considered in all dimensions and no samples are rejected, Gibbs sampling is a very efficient way of sampling highly dimensional distributions (as seen in 5). However, there is a weakness to the method, which restricts its use somewhat in weak lensing. As you can see from Figure 4.7, the size of the 'jumps' is determined by the width of the distribution along that axis. This is fine for mainly circular distributions like the one in the example. However, in highly correlated distributions such as the banana-shaped distribution of Figure 2.8 climbing up and down the banana one dimension at a time can be incredibly inefficient.

With this in mind, there are weaknesses to all MCMC and Gibbs-based methods that become highly important, especially when sampling the probability distribution of cosmological parameters. First, they all revolve around jumping methods that suffer from taking a minimal number of samples at low distributions (as there is such a low probability of those samples being accepted). At first, this is fine as we want the number of samples to be proportional to their distribution, but this can lead to samples becoming trapped within a region of high probability (local maxima) of the distribution is bimodal, but the maxima are considerably distant it can take a long time to escape one of the maxima. Secondly, ignoring the lower probabilities reveals an issue with one of our advantages. We initially said that these methods do not care if the distribution is normalised as we only care about the proportions, not the individual probability values. Once again, this is fine unless one wants to use the distribution for an untouched (within this thesis) advantage of Bayesian statistics, the ability to compare models. This is an aspect of analysis that requires the evidence and thus integrals of the entire distribution.

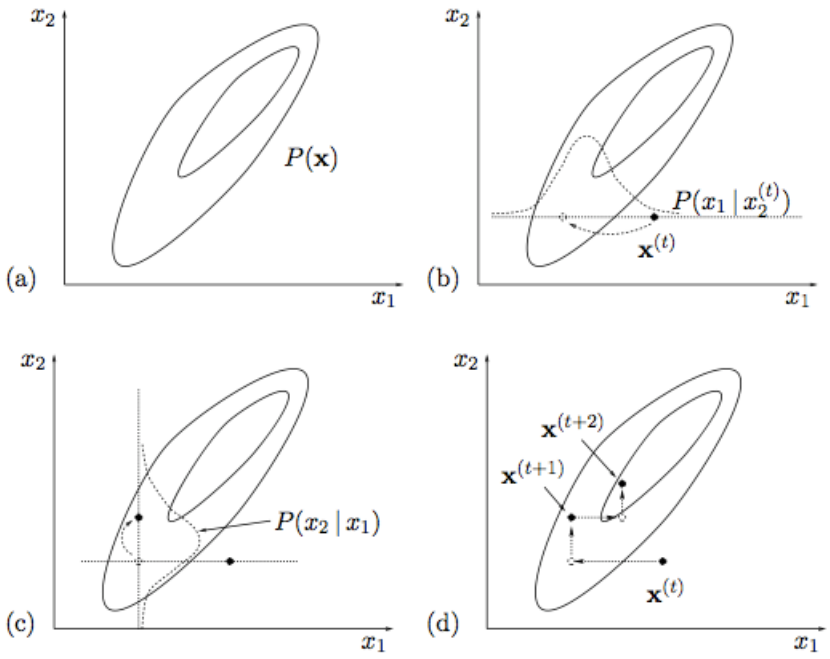


Figure 4.7: An illustration of the steps required in a 2-dimensional example of a Gibbs sampling procedure (Mackay, 1995).

Under these circumstances, nested sampling becomes a convenient approach (and the approach used for the cosmological analysis of Chapter 5 as well.). The exact algorithm is slightly more complicated than we have seen previously. If one wanted to compute the evidence seen within the Bayes theorem, one is looking at the volume (in parameter space) of the un-normalised distribution

$$\mathscr{Z} = \int \mathscr{L}(\theta)\pi(\theta)d\theta, \tag{4.15}$$

where $\mathscr{Z}, \mathscr{L}(\theta)$ and $\pi(\theta)$ are the evidence, likelihood and prior respectively. This is a multidimensional integral, so it can be difficult to compute. Instead, imagine a load of random points are taken along the distribution and contours are formed around lines of similar likelihood (see Figure 4.8 *left*). The prior volume contained within these contours is known as

$$X(\mathscr{L}) = \int_{\mathscr{L}(\theta)>\mathscr{L}} \pi(\theta)d\theta. \tag{4.16}$$

If we then integrate this volume $X(\mathscr{L})$ along the contours as one can see in Figure 4.8 *right* we extract the evidence $\mathscr{Z}$

$$\mathscr{Z} = \int_0^1 \mathscr{L}(\mathscr{X})dX. \tag{4.17}$$

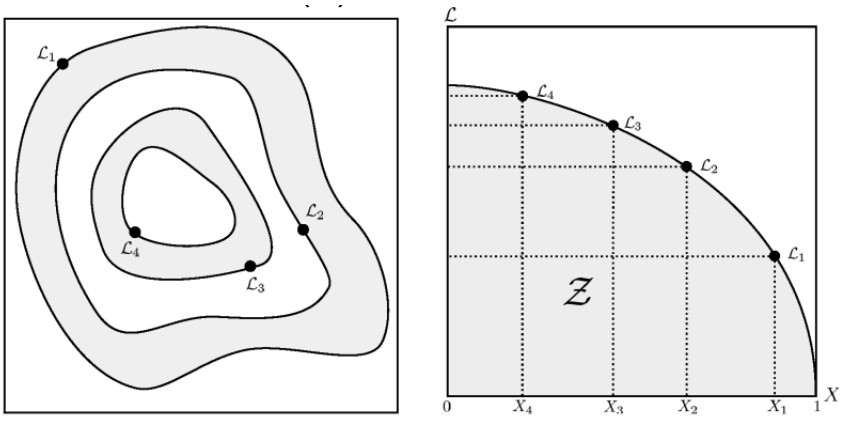


Figure 4.8: *left*: An illustration of the likelihood contours that are created around a probability distribution *right*: A plot of X against $\mathscr{L}(X)$ showing how the integral leads to the evidence $\mathscr{Z}$ (Pritchard, 2016).

4.2.2 BPZ

For the first direct implementation of Bayesian statistics into photometric redshift estimation, we do not need to worry about sampling methods due to the small parameter space in which our problem lies. As a reminder, 3.2.3 look at template-based methods of estimating redshift, showing huge promise due to their performance not being linked to spectroscopic data availability but instead based on physical theory. The main issue left lingering was that of template degeneracy. What can one do about the likelihood that functions of template-based methods are inherently poly-modal due to template degeneracy? One possible remedy comes from BPZ: Bayesian Photometric redshifts (Benitez, 2000).

Benitez realised that under the previous frequentist approach to template-based methods, simply increasing the number of templates to solve the template incompleteness issue only increased the colour/redshift degeneracies problem. This was due to all templates having equal status within the analysis, and thus likelihood functions on their own are not informative enough.

To take a fully Bayesian approach, one must evoke the Bayes theorem from equation 4.6 in which we want to find the posterior probability that an object has a curtain redshift z given its colour C and magnitude m_0:

$$p(z|C,m_0) = \frac{p(C|z)p(z|m_0)}{p(C)} \propto p(C|z)p(z|m_0). \tag{4.18}$$

The most important new segment is the inclusion of a prior $p(z|m_0)$. This allows us to 'sway' redshift estimations in the correct direction if the likelihood has very different redshifts with an equally high likelihood. The most obvious example would be the inclusion of a prior

that states that there should be a low probability of finding high redshift objects of high magnitude, as we may not have the practical capabilities of detecting the said object. The prior even allows us to take into account the fact that the magnitude observed m_0 is not necessarily the same as the true magnitude of the object $\hat{m}_0$ assuming the probability of change from true to observed is Gaussian gives the following prior

$$p(z|m_0) \propto \int d\hat{m}_0 p(\hat{m}_0) p(m_0|\hat{m}_0) p(z|\hat{m}_0) \tag{4.19}$$

The issue of degeneracy can now be truly attacked as different templates no longer have to be weighed equally. We can include that certain template types T would be less likely than others given a certain magnitude

$$p(z,T|m_0) = p(T|m_0) p(z|T,m_0). \tag{4.20}$$

The effect can be seen in Figure 4.9, of which $p(C|z,T)$ is the likelihood so that under a usual ML method, this object would be considered high redshift. However, due to a prior that favours all template types at lower redshifts, when one marginalises over all types, one reveals a posterior that favours a low redshift.

As one might expect, as the introduction of a prior is the main differentiator between BPZ and previous likelihood-based methods, the choice or creation of said prior is of much interest. In the situation where the prior is very vague, the original BPZ paper introduced a calibration method, assuming the underlining prior follows a particular analytic function with a set of free parameters λ. The fixing of this set of parameters will be governed using the available galaxy catalogue through the following estimation;

$$p(\lambda|C,m_0) \propto p(\lambda) p(C|m_0,\lambda) = p(\lambda) \prod_{j,j\neq i} p(C_j|m_{0j},\lambda), \tag{4.21}$$

then the posterior would be found through marginalising over λ

$$p(z|C,m_0) \propto \int d\lambda\, p(\lambda|C,m_0) \sum_T p(z,T|m_0,\lambda) p(C|z,T). \tag{4.22}$$

Benitez provided an example where the method was applied across the Hubble Deep Field North data (Williams and team, 1996). The underlining function was assumed to be exponential and low probability at high redshifts leading to the $p(T|m_0) = f_t \exp\{-k_t(m_0-20)\}$ for the spectral type prior and $p(z|T,M_0) \propto z^{\alpha_t} \exp\{-[z/z_{mt(m_0)}]\}$ were $z_{mt}(m_0)$ is the medium redshift given by $z_{mt}(m_0) = z_{0t} + k_{mt}(m_0-20)$ which leads to 11 free parameters

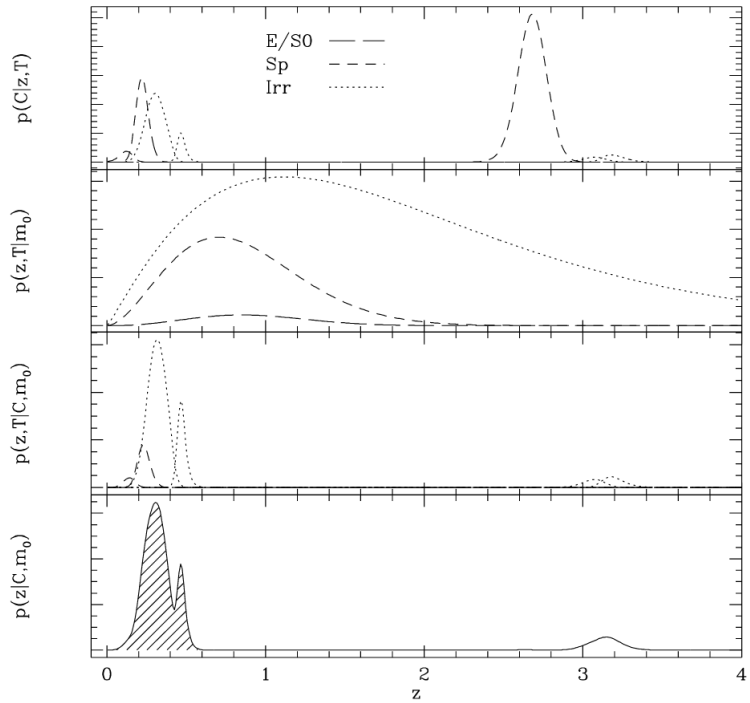


Figure 4.9: Example of how a probability distribution may change with the introduction of a prior choice, the first row shows the likelihood for each galaxy type, row 2: the prior, row 3: the posterior, row 4: the posterior marginalised over each galaxy type (Benitez, 2000).

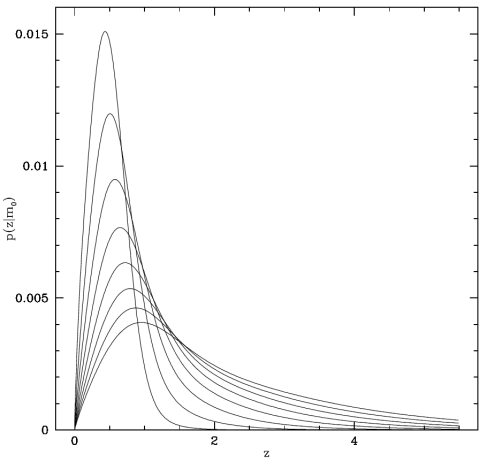


Figure 4.10: Prior in redshift and different given magnitudes $p(z|m_0)$ estimated through using the calibration method on the HDF-N data (Benitez, 2000).

$\lambda = \{\alpha_t, z_{0t}, k_{mt}, k_t\}$ these free parameters are found through the method mentioned above and a prior is found as shown in Figure 4.10.

The advantages of such a method are compelling. One can start with a very loose idea of what the prior should be and still gain a usable prior through the same galaxy sample. Furthermore, due to its rational nature, the resulting redshift function is smooth, and the effects of cosmic variance are reduced. There are, however, a few downfalls; the selection criteria of the galaxy catalogue may not be taken into account by the analytic form of the prior, and assumptions need to be made to constrain the fit of the free parameters (Feldmann et al., 2006).

To counter the limitations mentioned above, a new approach was suggested in which an iterative prior is used (Brodwin et al., 2006). In this method, they do not apply individual likelihoods of each galaxy to the prior but instead directly sum up the normalised likelihood functions L^n for each galaxy. This summed-up likelihood is the initial stage of the prior iterative method

$$[p(z,T|N(z))]^0 = \sum_i \mathscr{L}_i^n(z) = \sum_i \frac{\mathscr{L}_i(z)}{\int_0^\infty \mathscr{L}_i(z)dz}, \tag{4.23}$$

where $N(z)$ is the underlining redshift distribution of the sample. Using this prior, a first stage posterior is found, which is considered our new L^n and the iterative process is repeated until convergence. If this were spectroscopy, the first stage prior would be the correct redshift distribution (or at the very least a much more accurate estimate) but the photometry, particularly at the fainter magnitude, limits the derived first stage before containing artificial peaks. Continuing the iterative process, the bulk information is used to modify the likelihoods, which minimises degeneracy. One large assumption is that a prior is required for a

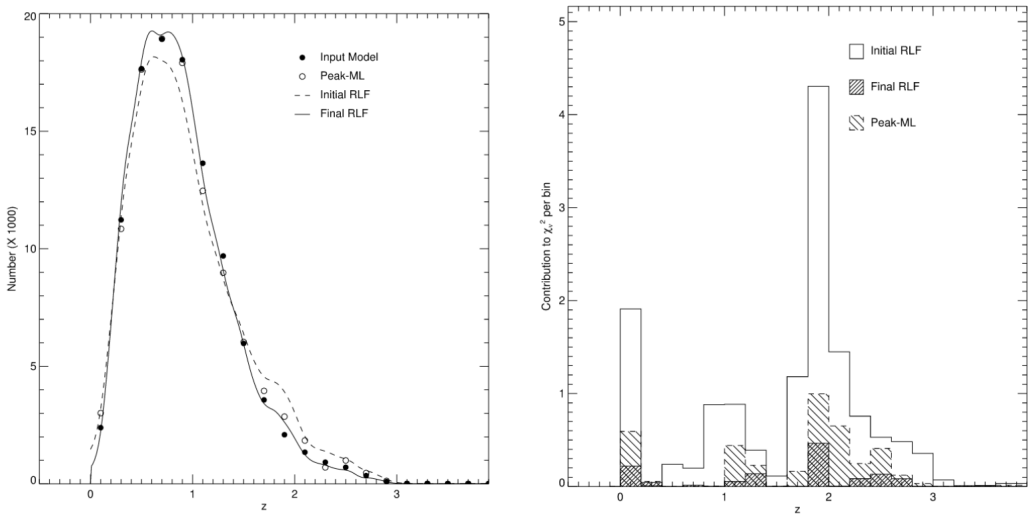


Figure 4.11: *left*:Results of a Monte Carlo simulation based on CNOC2 evolving luminosity functions (*solid circle*) well as the $N(z)$ distribution estimated by peak-ML (*open circles*), initial iteration (*dashed line*) and final iteration (*solid line*) *right*: The χ^2 contribution of the same N(z) distribution (Brodwin et al., 2006)

galaxy regardless of its type or magnitude. Thus individual photometric estimations are not necessarily improved only the complete statistical redshift distribution. The smoothing of the prior can be seen in Figure 4.11, where the initial prior shows unwanted peaks that reduce the correct peak. Directly comparing the total χ^2 give { peak-ML,initial-iterative,final-iterative } $= \{4.02, 14.5, 1.25\}$ showing the effectiveness of the method.

The advantage of this approach is that it can match over/under densities due to cosmic variance where for example, the smooth functions of BPZ would not pick up. However, with this said, the method no longer shows improvement to the initial prior at very low signal-to-noise ratios. It thus is recommended only to be used in the $0.2 \leq z \leq 1.3$ redshift range (With this said, Figure 4.11 recovered the correct N(z) outside of this suggested range, so it still has its uses).

Due to the success of this approach, it was adopted for the *Zurich Extragalactic Bayesian Redshift Analyser (ZEBRA)* (Feldmann et al., 2006) with the formulation extension of applying it across redshift *and* template leading to the following prior:

$$p_{\rm new}(z,t) = p_{\rm old}(z,t)\frac{1}{N_G}\sum_{i=1}^{N_G}\frac{\mathscr{L}_i(z,t)}{\sum_{z\prime,t\prime} p_{\rm old}(z\prime,t\prime)\mathscr{L}_i(z\prime,t\prime)}, \tag{4.24}$$

with N_G being the number of galaxies in the sample and the initial $p_{\rm old}(z,t)$ is a user-specific guess prior (usually flat). Other improvements include the fact that unreliable redshift estimations are excluded in the process. A 'bad fit' is any estimation with χ^2 values more

significant than the threshold $\chi^2_{0.99}$. In practice, this only excludes 1% of the data. The smoothing of the prior is also a noticeable improvement; due to a finite number of galaxies being observed, there are often high-frequency fluctuations in the observed $N(z)$, and these sharp features become more prominent with every iteration leading to a spiky prior. To account for this, the prior is smoothed after each iteration. This is achieved by smoothing the number counts with a Gaussian kernel of width $\sigma = max(\sigma_{ML}, \sigma_{\mathscr{L}}, \sigma_P)$ were σ_{ML} is the standard error of the maximum likelihood estimator, $\sigma_{\mathscr{L}}$ is the typical broadness of the likelihood functions and σ_P the characteristic scale due to finite Poisson sampling, most of which are found through Monte Carlo simulations. With all these improvements, studies suggest that ZEBRA and BPZ, when applied to the same samples out, perform their likelihood only (HYPERZ etc.) counterparts with the same templates (Abdalla et al., 2008).

4.3 Redshift Distributions

4.3.1 Stacked Likelihoods

With an understanding of how Bayesian statistics can be directly involved in producing photometric redshift distributions, it is worth returning to the overall goal of improving weak lensing analysis. For this, we must go back to the lensing efficiency equation, and the term of great interest to us is $n(\chi)_i$ or, more specifically, $n(z)_i$ as $n(\chi)_i d\chi = n(z)_i dz$. $n(z)_i$ refers to the redshift distributions of the i^{th} tomographic bins sources. As stated in chapter 2, we must group our observed sources into different groups to measure their correlations with each other. The way the sources are grouped does not matter, but if one was to group vaguely by redshift, one could potentially gather extra information by seeing how the lensing of galaxies (and thus the amount of matter the light had travelled through) changes with redshift. In this sense, BPZ is used directly. Even the most modern studies of weak lensing use the maximum posterior acquired by BPZ to place the sources into groups characterised by the maximum and minimum BPZ redshift estimations they would allow (Asgari et al., 2020). With that said, due to the probabilistic nature of BPZ, 'bleeding' can occur when galaxies are placed in entirely wrong bins. This can affect intrinsic alignment, which may use the ordering or sources to identify which ones can lens others. Due to this bleeding, it is essential to understand how much overlap between different bins, the same overlap described in our most crucial variable of interest, the redshift distribution itself $n(z)$.

The n refers to the number of sources between any particular $z-\delta, z+\delta$ interval so that the $n(z)$ is just an exemplary histogram of collected sources. Of curse, the histogram area can always be normalised to give a probabilistic value; what is the probability that a randomly

selected source within a tomographic bin lives within the redshift interval $z-\delta, z+\delta$? Note that this is a very different question from the ones being answered by BPZ and other photometric codes. We are no longer interested in individual galaxies but in the collective statistics of all the gathered sources. The traditional technique in this situation would be to reduce the photo-z PDFs down to delta functions $\delta(z,\hat{z})$ where $\hat{z}$ is a point estimate such as maximum posterior (Abruzzo and Haiman, 2019). This makes sense in terms of intuitively. We are forming a histogram, and a galaxy can only be at one point at any time. That said, such a method negates the work above, as so much information is stored within the probability distribution of the photo-z estimate. This is mainly due to the inherent error and uncertainties within the template-based methods. If these were spec-z estimations using this method would be entirely agreeable. Instead, the most common method across weak lensing has been to 'stack' the likelihood or posterior (Kelly et al., 2012)(Benjamin et al., 2013)

$$\hat{n}(z) = \frac{1}{N}\sum_{i}^{N} p(z|\hat{D}_i), \tag{4.25}$$

where N is the number of galaxies in the group. You are, in a sense taking a mean distribution by all the individual redshift distributions. Once again, there is some intuitive reasoning behind this choice. If one has a set of bimodal distributions of which the lower redshift peak is always highest, the delta method above would say the entire population is at the peak redshift, while the stack method says that 'on average you would expect some of the population to exist near the second peak and that they only appear to be lower redshift due to errors in observation. The stacked likelihood method then uses the entire distribution. Because of these benefits, the stacked method has been used successfully in weak lensing studies for years. Take Figure 4.12 which shows the estimated redshift distribution through stacked likelihoods doing a fairly good job compared to their spec-z 'true' distribution.

With that said, one can already see the failings that have led to the modern phasing out of stacked approaches to the extent to which modern weak lensing studies such as (Hildebrandt et al., 2017) KiDS-450 rejected the technique outright due to being "biased at a level that cannot be tolerated by contemporary and especially future cosmic shear measurements" (Newman et al., 2015). As we can see from Figure 4.12 stacking removes the ability to notice sharp features in the distribution because by averaging over all distributions, you are 'smoothing' any of their features. Secondly, the distribution begins to fail for the last tomographic bin, which could be linked to a higher redshift source probably being of higher magnitude and thus may be of much lower signal to noise leading to the catastrophic errors in distributions typical of template degeneracy. The issues with stacked likelihood are 2 fold. Firstly consider this practical example of why bias can develop; we have a set of galaxies

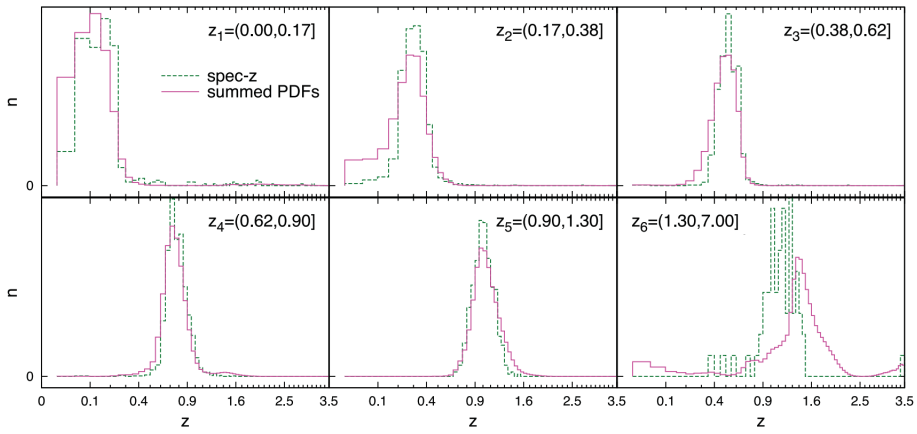


Figure 4.12: From the CFHTlens weak lensing analysis. Comparison of the stacked $n(z)$ posterior distribution for different tomographic bins against the 'true' distribution given through spectroscopic redshifts (Benjamin et al., 2013).

all of the low redshift $z = 0.1$ but with significant errors in their observations. For the sake of simplicity, we can limit ourselves to one template and say that the posterior distribution mimics a gaussian. The issue is that due to the nature of redshift, we know that $z > 0$ so the tail of the gaussian only develops on one side. The outcome of the stacked likelihood approach would be a gaussian distribution cut short at low redshifts and with a high redshift tail. Not only is this distribution not matching the actual 0.1 delta function, but the mean redshift of the distribution would bias to higher redshifts, and as we saw earlier, it is this mean redshift that can directly affect our lensing power spectrum results just as much as cosmological parameters. The second is that fundamentally the method is mathematically inconsistent (Hogg, 2012). If we examine equation 4.25 again, we can see that it almost resembles a marginalisation but with discreet instead of continuous variables. It looks as if we have 'marginalised' out the individual data of individual sources to produce the $n(z)$ of the whole survey. This is mathematical; one cannot integrate over a fixed value on the right side of conditional probability. Integrating over a free variable on the left side of a conditional (Malz, 2021) is only valid. So how is it that stacked likelihoods were so prominent for so long? Malz (2021) examined this exact problem and concluded that there are 2 situations in which stacked posteriors recover the exact redshift distribution:

1. when ones have perfect informative data
2. when one has a perfectly informative prior

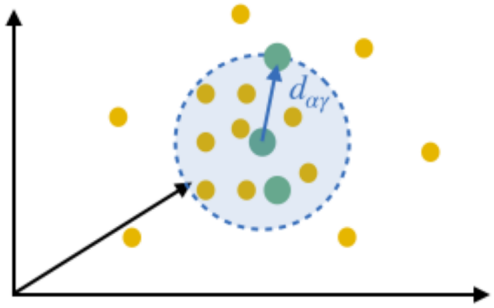


Figure 4.13: Illustration of the nearest neighbours method to which the yellow dots represent photometric samples, the green dots the nearest spectroscopic neighbours of our chosen galaxy and blue the volume of radius $d_{\alpha\gamma}$ of who is density must be recalibrated. In this case, the magnitude space is 3 dimensional, but in reality, the number of dimensions would be the number of filters (Gomes et al., 2018).

2 conditions that were only weakly violated in the past. Nevertheless, as we move towards deeper surveys of decreasing signal to noise, these conditions will continue to be violated to a larger extent.

4.3.2 Spectroscopic Based Methods

With stacked likelihoods/posteriors clearly show a lack of ability to work under the conditions of modern weak lensing surveys. The commonality to which photometric redshifts were used directly for redshift distribution estimations dwindled to such an extent that their use by weak lensing surveys has been close to abandoned with the likes of the Dark Energy Survey (Hoyle et al., 2018), the Kilo Degree Survey (Hildebrandt et al., 2017) only using photometric redshifts for the original grouping but not the distribution estimation at all. For that, modern surveys have returned to using spectroscopic data.

Although the first use can be seen in an application directly to weak lensing statistics in the work of Bonnett et al. (2016), the method was developed and described in its entirety in the works of Lima et al. (2008). Known as the weighted direct calibration method (DIR), the method seeks to calibrate the much more accurate redshift distribution of a spectroscopic survey using the magnitude data of the photometric data. In more detail, the method begins by taking photometric and spectroscopic surveys. In the case of the KiDS-450 photometric survey, a combination of multiple spectroscopic surveys was used, including the Cosmic evolution survey (COSMOS) (Scoville et al., 2007), specifically a non-public catalogue provided by the zCOSMOS team (Lilly et al., 2009), the DEEP2 catalogue (Newman et al., 2013b), and a compilation of spec-z from Chandra Deep Field South (CDFS) (Vaccari et al., 2012) released by ESO. A redshift distribution $p_T(z)$ (where the $_T$ denotes the training

sample) is known for the spectroscopic sample, but it has not yet been calibrated. Both surveys are plotted separately on a magnitude hypersurface based on their filter bin values. The goal is to reweigh the spectroscopic galaxies so that their magnitude surface density matches the photometric one. This is achieved through the nearest neighbours approach as illustrated in Figure 4.13; one by one, a single spectroscopic galaxy is selected, and the euclidian distance $d_{\alpha\gamma}$ between that galaxy (α) and every other galaxy (γ) is calculated $d_{\alpha\gamma}$

$$(d_{\alpha\gamma})^2 = \sum_{i}^{N_f} (m_\alpha^i, m_\gamma^i)^2, \tag{4.26}$$

where N_f is the number of filters. A hyper-volume of radius $d_{\alpha N_{\rm nei}}$ is formed as the distance between the selected galaxy and its $N_{\rm nei}^{\rm th}$ closest neighbour (a fixed value chosen by the user). We now go to the same magnitude-based location of the chosen spectroscopic galaxy but on the photometric magnitude surface and count the number of galaxies that are fully within its volume so that $d_{\alpha\gamma} < d_{\alpha N_{\rm nei}}$. With the number of galaxies within the volume on both surfaces known, one can construct a weighting W_α

$$W_\alpha = \frac{1}{N_{p,\rm tot}} \frac{N_p(m_\alpha)}{N_T(m_\alpha)}, \tag{4.27}$$

where $N_{p,\rm tot}$ and $N_p(m_\alpha)$ are the total number of photometric galaxies and the number of photometric galaxies within the distance $d_{\alpha N_{nei}}$ around the magnitude of the chosen spectroscopic galaxy m_α. While $N_p(m_\alpha) = N_{\rm nei}$ is the fixed number of nearest neighbours around the α galaxy. After the weight has been found for every spectroscopic galaxy, the photometric redshift of the survey $p_T(z)$ is recalculated by counting the number of galaxies within each histogram bin $[z-\delta, z+\delta]$. However, this time instead of having their discrete values of 1 and 0 (they are there or not), they use their weighted value if they are there so that spectroscopic galaxies whose magnitudes are not well represented within the photometric survey have less impact on the overall distribution.

As with many methods, weaknesses and possible sources of error come through its ad-hoc choices. In this case, the choice of what $N_{\rm nei}$ to use. Ideally, $N_{\rm nei}$ needs to be small enough to preserve locality, smaller than the typical size to which $N_p(m)$ varies; otherwise, one would be smoothing over differences in the distribution. However, the issue of locality can have quite an effect. For example, imagine a galaxy with 4 other galaxies closely clustered in magnitude space. When one wishes to use a $N_{\rm nei} = 5$ nearest neighbours approach, the hypersphere volume will be huge, incorporating the fifth galaxy much further away in magnitude space. When this vast volume is replicated in photometric space, it may collect many photometric galaxies under its volume, meaning these galaxies that are already sharply

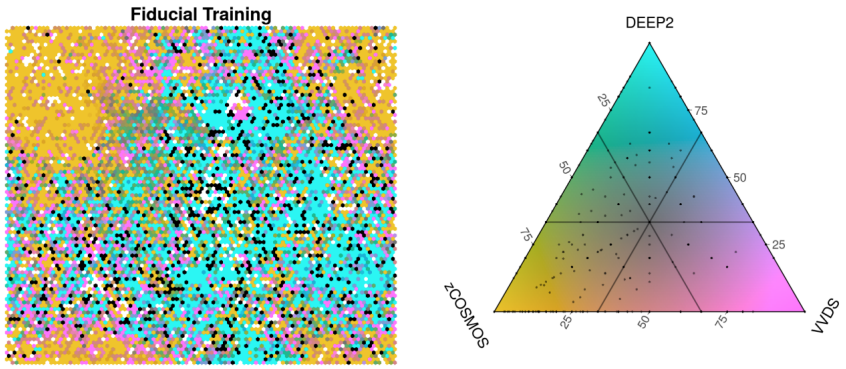


Figure 4.14: *left*: An example of the self-organised maps used by KiDs-1000 where the cells have been fictitiously coloured based on which spectroscopy galaxies from which survey are contained within them. All black cells represent cells containing only no spectroscopic galaxies *right*: The survey-based colour scheme (Wright et al., 2020).

clustered in one magnitude region will be considered up weighted, giving a false magnitude-redshift relation. With that said, choose a number a N_{nei} to small, and once it is susceptible to large amounts of shot noise, random noise would appear statistically significant to the distribution. Another issue of high importance is that this method relies on the spectroscopic survey being a good representation of the photometric survey. If it is not, then regions of the photometric magnitude space will be completely ignored and not considered in the final calibration. Plus, photometric galaxies with completely different redshifts to their spectroscopic magnitude neighbour may influence weightings. In reality, they should not be involved in the calibration of theta galaxy at all (or, more importantly, that spectroscopic galaxy should not be a representative of that photometric galaxy).

With this in mind, a new method was developed to tackle these issues systematically and is the current state-of-the-art method used by KiDS for their most recent weak lensing surveys (Hildebrandt et al., 2020). On the surface, both methods seem incredibly similar, taking a spectroscopic survey and reweighting the individual galaxies to match the magnitude space of a photometric survey. The difference comes in how the magnitude of space is constructed. Instead of using a N_f-dimensional field, a 2-D Self-organised map is used (the same maps as seen in Figure 3.9). The self-organised map was formed using the spectroscopic and photometric data, as shown in Figure 4.14. Spectroscopic galaxies are reweighed so that every cell contains the same number of photometric and spectroscopic galaxies.

Already this looks to address the first issue, clusters of galaxies will no longer be linked to an outlier because of a chosen N_{nei} instead, galaxies will only be reweighed based on other galaxies within their cell, ensuring that all galaxies have similar magnitude characteristics. In terms of the issue of the spectroscopic survey not representing the photometric survey

accurately enough, a culling process takes place in which any photometric galaxy placed in a cell containing no spectroscopic galaxies (i.e. black cells in Figure 4.14) are removed from the entire weak lensing analysis so that all relevant magnitude spaces are covered. A similar culling takes place by removing all photometric galaxies whose redshifts differ too much from those of the spectroscopic galaxies. This is decided by seeing the differences between the estimated z_{spec} and the mean photometric redshift of the photometric galaxies. In this sense, galaxies cannot be reweighted by galaxies in completely different redshift spaces.

The SOM approach marks the current best method for constructing redshift distributions, but through its successes, it is again leaning on resources and technologies one wishes to avoid for more profound surveys. For starters, relying on spectroscopic data is limited by how well they cover the photometric surveys in magnitude space. Even the KiDS-1000 saw their number density of photometric galaxies per square arcmin fall from 7.4 in their DIR approach to 6.2 approach due to this culling (Hildebrandt et al., 2020). These galaxies are no longer usable for where they are most needed, the weak lensing analysis. This fall will only continue as surveys become more profound and may become inefficient going forward. Secondly, they rely on machine learning techniques such as self-organising maps of which ad-hoc changes to cell size may have unknown ramifications to the redshift distribution, let alone the biases caused by using specific spectroscopic survey training data.

With this in mind, as successful as the above methods are, we are still left wanting a genuinely Bayesian method of inferring a redshift distribution with correctly propagated uncertainties through the use of the photometric data we have, without the inclusion of spectroscopic data and machine learning.

4.3.3 Principled Approaches

We return to constructing an appropriate Bayesian model for estimating the redshift distribution of a weak lensing survey through photometric data alone. To do we consider the work of Malz and Hogg (2020) who takes the following steps to allow for such parameter inference, construct a forward model and thus the likelihood, specify prior beliefs, and compute the posterior (Alsing, 2017). By forwarding the model, we are referring to the order in which Bayesian Hierarchical Models are constructed in which the starting point is parameters, and the ending is the observed data. Named Cosmological Hierarchical Inference with Probabilistic Photometric Redshifts (CHIPPR) (Malz and Hogg, 2020) the model is constructed in simple form in Figure 4.15 and expressed through 3 steps

1. A redshift distribution that has been parameterised so that $f(z;\phi) = n(z)$ where ϕ is a set of hyperparameters

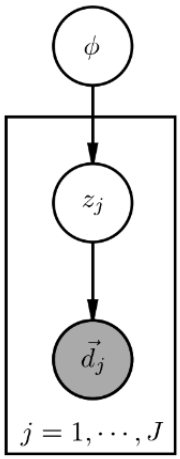


Figure 4.15: A direct acyclic graph of the CHIPPR model consisting of the ϕ the hyperparameters that describe the redshift distribution $n(z)$, a set of redshifts z_j and their corresponding photometric data $\vec{d}_j$ for a set of J galaxies (Malz and Hogg, 2020).

2. A set of J galaxies have there redshifts z_j independently drawn from the redshift distribution so that $z_j \sim p(z|\phi) = f(z;\phi)$
3. observed photometric data $\vec{d}_j$ is gathered drawn from a likelihood function $p(\vec{d}_j|z_j)$

Some other assumptions are worth delving into in more detail. For example, the model presumes there is a higher dimensional probability space $p(z,\vec{d})$ such that the $n(z)$ can be described as an integral

$$n(z) = \int p(z,\vec{d})d\vec{d}. \tag{4.28}$$

This particular distribution does not have to be 1-1 as galaxies of the same redshift can have different observed photometry (such as being different galaxy types), just as galaxies of different redshifts can have the same photometry (due to template degeneracy). Furthermore, we do not care about the galaxies' redshifts under this framework. We only care about the underlying distribution (and thus the parameters that make it) to which the redshifts are drawn so that the posterior of which we are trying to infer is $p(\phi|\vec{d}_j)$ (notice no z_j).In this simple form, we have missed out on some of the subtle points that make BHMs, such as all the latent variables underlying the relationships between the main steps. Returning to our idea of the joint probability $p(z,\vec{d})$ if one was to estimate a single galaxies redshift they are essentially replicating equation 4.28 for the special case where $\vec{d} = d$, one can even bring in our hyperparameters as the true joint probability if which we are marginalising over is $p(z,\vec{d}|\phi^\dagger)$ where $\phi^\dagger$ is the true redshift distribution parameters. In reality, $\phi^\dagger$ is unknown and yet an underlying distribution is assumed by every redshift estimation method we have seen so.

By the construction of a model, either machine learnt or template-based one has constructed a joint probability between redshift and photometry that can be described as $p(z,\vec{d}|\phi^*)$ where ϕ^* is the parameters of the underlying distribution as dictated by the chosen model. This may seem like an unnecessary step but is a direct consequence of both equation 4.28 and the rewriting of $n(z)$ as $f(z;\phi) = n(z)$. If a chosen model implies a relationship between redshift and photometry, then a distribution $p(z|\phi^*)$ can always be found through marginalisation. Thus, an underlying distribution as dictated by the chosen model must exist. This underlying distribution is known as the *implicit prior* and also is rarely known so that individual redshift posteriors are usually conditional to it $p(z|\vec{d},\phi^*)$. An explicit prior $p(\phi)$ which enforces our prior beliefs not on the parameters of the underlying distribution of the model ϕ^* but on the parameters of the distribution of the survey ϕ (the parameter we care about). Such priors may be direct in prescribing some regions of z to be denser than others, or they may be less direct such as prescribing smoothness to the whole distribution. Through the terms mentioned above, one can construct a full posterior (for the full derivation, please see the Appendix of the work of Malz and Hogg (2020))

$$p(\phi|\vec{d}_j) \propto p(\phi) \int \prod_j^J \frac{p(z|\vec{d},\phi^*)}{p(z|\phi)} p(z|\phi^*) dz, \tag{4.29}$$

which is sampled for the truth distribution. The model can be incredibly effective and demonstrates some floors with previous approaches of using single redshift estimations or stacked methods. Take Figure 4.16 as an example. On the left, we can see the simulated data. In this case, we have the perfect scenario where the implicit prior is uniform (thus, it is a model in which there is no bias for any particular redshift) but the photometric redshifts do suffer from scattering and catastrophic error when compared to their spectroscopic counterparts. If we compare different methods for obtaining the redshift distribution as we see from *right*, the CHIPPR samples maintain the true distribution, whereas the other approaches suffer from their known weaknesses. Taking the maximum posterior from each z_{photo} (MODE-yellow) reveals it to be completely dictated by the outliers as it does not have access to the rest of the distribution. Whereas the stack posterior (Stacked-red) full prey to its large high redshift tails

The discrepancy becomes even more evident if we consider the expected results within a template-based framework (to be clear, the results can be replicated with very similar results using a machine learning framework as well). By this, we mean the simulated data now have an implicit prior that biases specific redshifts, as seen in Figure 4.17 *left*. This bias can be from template degeneracy in which certain z_{spec} produce different data that all point towards the same z_{photo} as well as the expected level of scattering and catastrophic error. Under these conditions, CHIPPR is still incredibly robust, whereas MODE and Stacked are

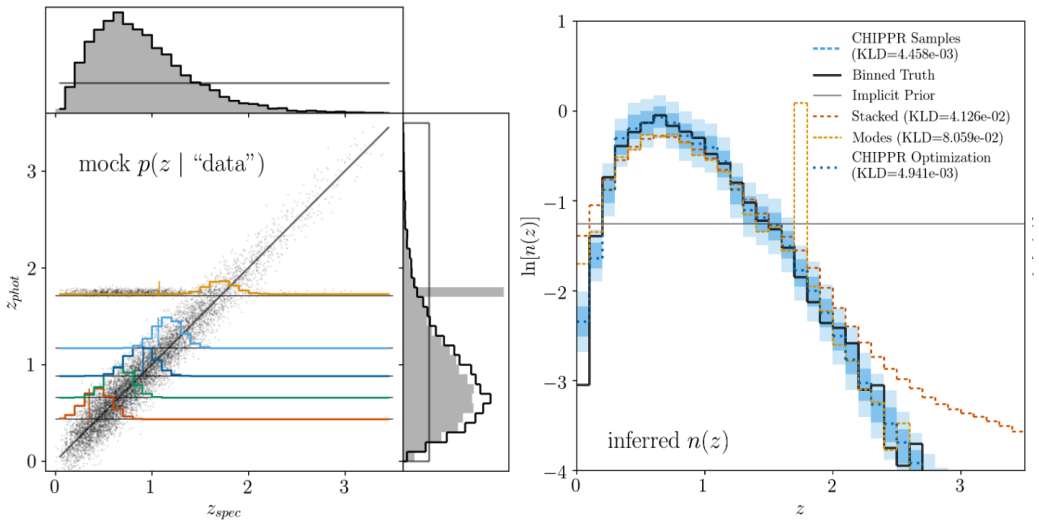


Figure 4.16: *left*: Example of simulated data plotting their true spectroscopic redshifts against their expected photometric redshifts, in this case, there is a non-biased uniform implicit prior (grey line in histograms) but with a level of scattering and catastrophic error so that they photometric distribution (grey) differs from the spectroscopic distribution (bold black line). *right*: A comparison of the posterior samples of CHIPPR (blue) against the true distribution (black) as well as other distribution techniques (Malz and Hogg, 2020).

wholly dictated by their model, i.e. the implicit prior. This is quite damming in terms of how helpful said methods can be if the estimation model choice can have such an effect.

As effective as these results seem, our analysis has glaring omissions. The first is relatively minor, which is that the computing posterior is computationally costly (through being both a product and an integral), and thus methods of optimisation are left wanted by Malz and Hogg (2020). The second major weakness is that all the simulations above did not simulate photometric data. Instead, it implied the redshift estimation step has already taken place, and what one has is a set of redshift posteriors $p(z_j|\vec{d}_j, \phi^*)$. This is to ensure ϕ^* is known and fixed so one can implement the implicit prior into equation 4.29. In reality, knowing the implicit prior can be incredibly difficult, yet it is imperative. For example, consider Figure 4.18, in this case, the same data has been used from Figure 4.17, but in this case, the implicit prior used in the analysis is uniform and thus incorrect (grey line). Under these conditions, CHIPPR fails, performing as bad if not worse than its MODE and Stacked counterparts. Thus for this method to be valid, methods must be developed that allow one to access the implicit prior. As said, this may be incredibly difficult. For template-based methods, at least we know the model which was constructed (we chose the templates), but this does not mean one has access to the joint $p(z, \mathrm{SED})$ posterior, let alone a method to marginalise it and find a set of hyperparameters. It may be even worse for machine learning estimation methods due to the

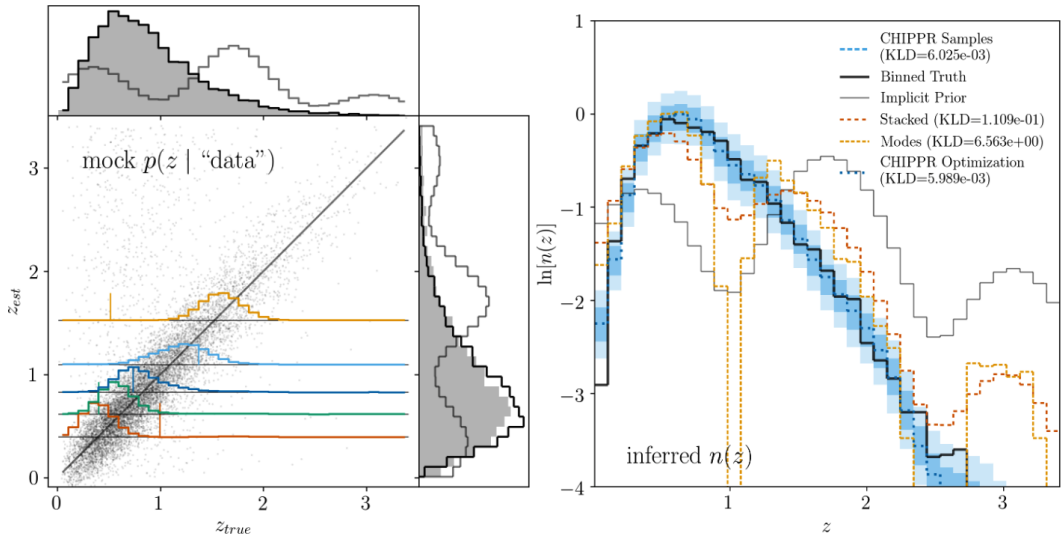


Figure 4.17: *left*: Example of simulated data plotting their true spectroscopic redshifts against their expected photometric redshifts, in this case, there is a template-based implicit prior (grey line in histograms) but with a level of scatter and catastrophic error so that they photometric distribution (grey) differs from the spectroscopic distribution (bold black line). *right*: A comparison of the posterior samples of CHIPPR (blue) against the true distribution (black) as well as other distribution techniques (Malz and Hogg, 2020).

black box nature of which data leads to estimation. CHIPPR may be the most principled and mathematically correct approach to understanding distributions, but its current form is unusable due to the need for a correct implicit prior.

4.4 Summary

The above chapter considered a completely different way of looking at probability. Instead of the frequency at which an event would take place (frequentist), it is the belief one has that an event will occur (Bayesian). This framework is much more approachable for cosmological science due to the single set of observables available to us (the universe) instead of the many millions of repeated experiments that can take place in other fields such as particle physics. Furthermore, the Bayesian approach allows us to consider priors, the information we know before the experiment has taken place that may contribute towards the final probability. This prior differentiates the probability of Bayesian statistics (the posterior) from that of frequentist statistics (the likelihood). Even so, great care is needed in choosing said prior due to its influence on the final result.

Due to the computation involved in Bayesian statistics, new methods of the sampling distribution (as opposed to evaluating every parameter value) were invented to dramatically

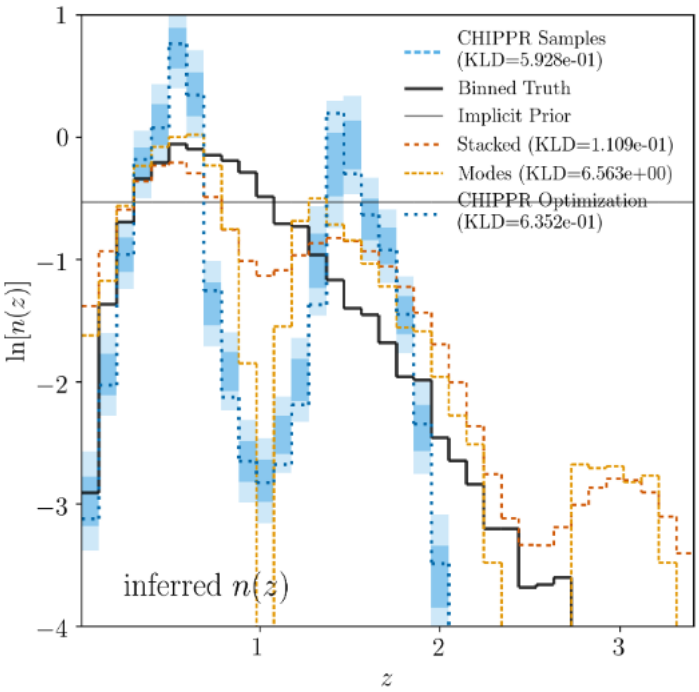


Figure 4.18: A comparison of the posterior samples of CHIPPR (blue) against the true distribution (black) as well as other distribution techniques, in this case, the same data from Figure 4.17 is used but an incorrect uniform implicit prior in the analysis (Malz and Hogg, 2020).

lower the number of single posterior evaluations required to analyse the whole distribution statistically. The Bayesian approach was then applied to template-based redshifts in the form of BPZ, showing the power that adding an informative prior does to help the issue of template degeneracy.

Lastly, we considered the true goal of the thesis, which is to obtain redshift distributions of the entire weak lensing survey as opposed to individual estimates so that one can use the $n(z)$ when computing the shear (or convergence) power spectrum and thus relate to the cosmological parameters involved in the matter power spectrum. Until recently, the most common methods have been to use the posterior point estimates or to average across all the distributions (stacking), but in recent years this is obsolete due to the reduction in signal to noise found in deeper weak lensing surveys. Instead, spectroscopic calibration methods have taken favour in the form of DIR and self-organising maps. Although practical, both methods rely on spectroscopic data that correctly covers the photometric data. If said coverage cannot be guaranteed, one may lead to incorrect calibrations (as for DIR) of significant reductions in usable data (self-organising maps). With this in mind, we returned to the Bayesian approach to solve the problem of redshift distributions based only on the photometric data. In this case, the most principled way was constricted by Malz and Hogg (2020) showing great promise compared to its stacked counterparts. However, due to its requirements for an implicit prior, the method is unusable in practical terms until methods of obtaining the implicit prior are developed. Instead, we will consider a different approach that still relies only on photometric data, exists as a Bayesian Hierarchical Model, but most importantly, through the work of this thesis is shown to work under practical circumstances, such as with the data of the Kilo Degree Survey to infer not only $n(z)$'s but complete weak lensing analysis in the form of S_8 inference.

Chapter 5

Algorithmic Development

Inference of cosmological parameters has become the primary goal of extensive galactic surveys. One can test cosmological models and constrain parameters within these models to agree with observational data from said surveys. One such observation would be the cosmic shear map; the contortion of the image of galaxies we see due to weak gravitational lensing, the bending of light in our line of sight by large-scale structure (Kilbinger, 2014).

As seen in Chapter 2, matching this map back to cosmological parameters requires an integral along the line of sight weighted by the source galaxy distribution $n(z)$. Thus the redshift distribution of galaxies (a crude estimation of the distance a galaxy is from us) within the survey is of utmost importance (Mandelbaum, 2017) and some methods to infer it were covered in 4.3. At first order the convergence power spectrum P_κ has comparable sensitivity to both cosmological parameters and median redshift $\bar{z}$ (Bernardeau et al., 1996; Huterer et al., 2006; Jain and Seljak, 1997). The accuracy of these distributions is critical in avoiding significant biases in final results. The demand for accurate $\bar{z}$ is exacerbated by using tomographic bins, the grouping of different galaxies to test correlation, usually by redshift. To deal with the vast amount of data required, most surveys rely on photometric redshifts, the measurement of redshifts through a small number of filters as opposed to entire spectra. Methods have been developed to deal with accurate estimations of low-resolution data, such as template fitting (Walcher et al., 2010) and, more specifically, BPZ (Benitez, 2000). Template fitting is the matching of photometric data to that of spectra of known redshift weighted by the appropriate filters.

As accurate as these approaches give estimations and posteriors to individual galaxies, they do not translate to developing usable and accurate $n(z)$ for an entire survey. Previous surveys such as CFHTLenS and DES used the stacking of individual redshift posterior distributions (Bonnett, 2015; Hildebrandt et al., 2012) (see 4.3.1). However, this approach magnifies the scatter of the underlying $n(z)$ and may not be the most suitable way to infer the distribution

(Hikage et al., 2019; Padmanabhan et al., 2005). This may make biases at a level intolerable for current and future surveys (Choi et al., 2016; Newman et al., 2015). It also does not give uncertainties in its $n(z)$ values.

Current surveys have attempted to alleviate this problem by using a complementary spectroscopic survey (see 4.3.2). One such method is the weighted direct calibration, DIR. First proposed by Lima et al. (2008), the method takes the individual redshift estimations gathered from the more accurate spectroscopic redshift estimations and calibrates the final $n(z)$ by re-weighting the spectroscopic redshift estimations so that the spectroscopic survey matches the broader but much less accurate photometric survey in magnitude-band space. This method is accurate and gives uncertainties in their estimations through bootstrap samples of said surveys. Unfortunately, this accuracy relies on an over-lapping spectroscopic survey (Joudaki et al., 2020), a luxury one will not have as photometric surveys become deeper in redshift and broader in scope (Euclid Collaboration et al., 2019; LSST Science Collaboration et al., 2009), both an issue of note for recent analysis (Joudaki et al., 2020). Furthermore, uncertainties are produced, but samples cannot be taken in a strictly Bayesian sense. Cosmic shear analysis from the first year data of the Hyper Suprime-Cam (Hikage et al., 2019) used a similar approach to DIR for gaining redshift distributions. The main difference is that overlapping spectroscopic surveys to HSC (Liske et al., 2015) do not have galaxy numbers large enough for the reweighing process to represent the properties of the photometric sources accurately. Alternatively, the HSC uses the COSMOS 30-Band Photo-z catalogue (Ilbert et al., 2008) as its accurate redshift estimates to be reweighed, but with this comes the extended uncertainties that come with base photometric data, be it with a significantly larger amount of bands when compared to KiDS. Alternative techniques for inferring the $n(z)$ redshift distributions are the CC and BOR methods, developed by Newman (2008) and Bordoloi et al. (2010) respectively. The CC method uses the cross-correlation functions between photometric and spectroscopic objects to find the photometric redshift distributions. On the other hand, Bordoloi et al. (2010) proposed the BOR method, which is strictly a recalibration of the posterior probability of each galaxy estimated using BPZ. A method gaining appropriate $n(z)$ distributions and uncertainties from photometric data alone is paramount to future weak lensing cosmology.

Instead of spectroscopic methods, Bayesian approaches were considered to get uncertainties and principled estimations for the redshift distributions. With that said, previous attempts (see 4.3.3) have been effective but unusable in a practical sense. With that in mind, below, we consider a different method to obtain their $n(z)'s$. We expand on the previous work by Leistedt et al. (2016), who presented a Bayesian Hierarchical Method (BHM) to infer both the redshift distributions and individual redshifts from photometric galaxy catalogues. The

methodology of said method is elaborated on below, with extensions included to make the method best suited to the tomographic bin nature of weak lensing data. Simulated data is created and used to validate the method and then applied to actual data to show its practical use. We then apply the redshift distributions to an actual cosmological data analysis problem and infer cosmological parameters whilst marginalising the nuisance parameters. Finally, the results are presented and discussed over the challenges associated with inferring $n(z)$ redshift distributions.

5.1 Method

5.1.1 Background Theory

The crux of the Bayesian approach is to consider the entire distribution of galaxies as a set of coefficients $\{f_{ijk}\}$ where ijk are the index for the template, redshift and magnitude bins, respectively. This model allows the survey to be represented as a 3-D histogram that approximates $p(t,z,m)$ such that

$$f_{ijk} = \int_{z_{j,\min}}^{z_{j,\max}} \int_{m_{k,\min}}^{m_{k,\max}} p(t_i,z,m) dz dm. \tag{5.1}$$

In the real use case, one would sum over the j,k cells to reveal the redshift distribution f_i. Each coefficient is now considered and individual parameters are inferred, the prior of which is uniform subject to constraints $0 < f_{ijk} < 1$ and $\sum_{ijk} f_{ijk} = 1$ which is a Dirichlet distribution.

The advantages of such a distribution are the ability to capture inter-bin correlations, which are vital for parameter inference. Furthermore, secondary redshift peaks usually associated with unwanted degeneracies in template methods are also suppressed -to an extent- as an additional advantage. Most importantly, the conditional posterior distribution of a Dirichlet is easily sampled by allowing the use of a 2-step Gibbs sampler between the individual $\{t_g, z_g, m_g\}$ and population $\{f_{ijk}\}$ parameters, respectively. The individual components are expressed through the hierarchical forward model in Figure 5.1.

Weak lensing analysis does not rely on the template, redshift and magnitudes of each galaxy but instead requires the distribution of the whole population. Under this requirement, one can speed the process using memory, and the speed-efficient version of the Gibbs sampler was suggested in the original paper. Moreover, in this form, one no longer requires the likelihood of each galaxy having any values for z,m,t but instead the likelihood that they

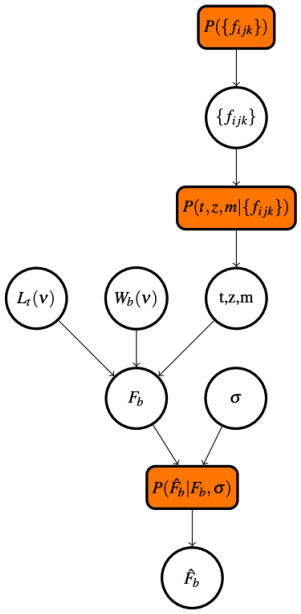


Figure 5.1: Hierarchical forward model for the the observed fluxes $\hat{F}_b$ of a survey of galaxies within a given band b where $W_b(\nu)$ is the b^{th} filter and $L_t(\nu)$ the spectral template of type t.$\{f_{ijk}\}$ is the distribution given the redshift, magnitude and template bins (i,j,k respectively), F_b the actual flux and σ the error on this flux.

will be within an $\{i, j, k\}$ bin from the 3D histogram

$$p(\{\hat{F}_b\}_g|ijk) = \int_{z_{j,\min}}^{z_{j,\max}} \int_{m_{k,\min}}^{m_{k,\max}} p(\{\hat{F}_b\}_g|t_i, z, m)dzdm, \tag{5.2}$$

this allows the writing of a simplified joint distribution,

$$p(ijk, \{f_{ijk}\}|\{\hat{F}_b\}_g) \propto p(\{f_{ijk}\}) \prod_{g=1}^{N_{\text{gal}}} p(\{\hat{F}_b\}_g|ijk)p(ijk|\{f_{ijk}\}). \tag{5.3}$$

The Gibbs sampling procedure then simplifies; given a sample of $\{f_{ijk}\}$, number counts for the number of galaxies in each point in the histogram n_{ijk} are drawn by drawing bin location samples from there likelihood function

$$p(ijk|\{f_{ijk}\}, \{\hat{F}_b\}_g) = f_{ijk} \times p(\{\hat{F}_b\}_g|ijk). \tag{5.4}$$

Using this n_{ijk} one draws samples for f_{ijk} following a multinominal distribution,

$$p(\{f_{ijk}\}|\{n_{ijk}\}) = (N_{gal} + N_t N_z N_m - 1)!\,\delta_D(1 - \sum_{ijk} f_{ijk})) \prod_{i=1}^{N_t}\prod_{j=1}^{N_z}\prod_{k=1}^{N_m} \frac{\Theta(f_{ijk}) f_{ijk}^{n_{ijk}}}{n_{ijk}!}. \tag{5.5}$$

The memory efficiency comes from the input being a likelihood distribution of much broader bins. Furthermore, there is no need for $N_{\rm gal}$ posterior distributions per sample as only the overall distribution is required. A necessity when dealing with large numbers of galaxies $N_{\rm gal} > 100000$, as with extensive photometric surveys. The individual parameters are discarded through this method, and template and magnitude parameters are marginalised to provide samples for $n(z)$.

The choice of a Dirichlet distribution can be considered in slightly more detail. For example, in its most traditional form, a Dirichlet distribution $\mathrm{Dir}(\theta|\alpha)$ can be expressed through Beta functions

$$\mathrm{Dir}(\theta|\alpha) = \frac{1}{\mathrm{Beta}(\alpha)} \prod_{i}^{K} \theta_i^{\alpha_i - 1}, \tag{5.6}$$

where for the context of work, θ, α is the probability and number count of an $i^{\rm th}$ bin and K is the total number of bins. Note that form is 1 dimensional and may seem different to our 3 dimensional f_{ijk}; however, one can always reduce the dimensions as a Dirichlet has no context to the physical nature of the bins, increasing computational efficiency. In one case, see its as a generalisation of a Beta function. Beta functions $\mathrm{Beta}(\alpha, \beta)$ intuitively give the probability of a binomial function having a certain success probability when the results of a repeated experiment give $\alpha - 1$ successes and $\beta - 1$ failures, as shown in Figure 5.2. A clear example is given by $Beta(5,5)$ which corresponds to 4 surceases and 4 failures, thus given distribution peaking at a success rate of 0.5, a peak that would only become narrower with more experiments.

Beta distributions can then be used as conjugate priors binomial distributions allowing for easy inference to 2D outcomes. Plus, the diversity of shapes (bell, U-shape, gaussian-like) makes them very useful for modelling real measurements. However, in the case of our redshift distributions, we have > 2 bins and thus > 2 outcomes. It is not a binomial but a multinomial distribution to which a Dirichlet is a conjugate prior. To visualise the distribution consider a random number generator that can only produce a 1,2 or 3 (we limited ourselves to 3 to allow for intuitive visualisation). We want to gather the probability associated with the generator producing each number, and for that, we observe a set of numbers produced, in the case we θ being the probability of each number and $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ being how many

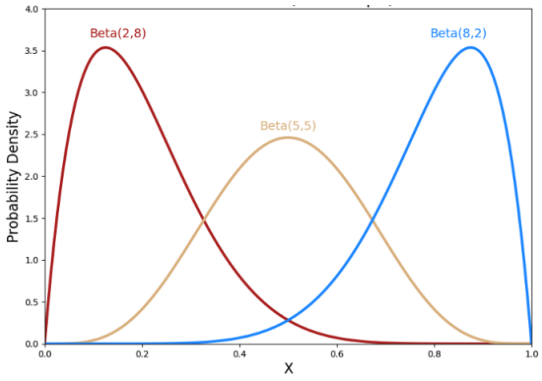


Figure 5.2: The beta distribution for varying values of α and β (Kim, 2020).

of each number (-1) was produced by the generator. The different Dirichlet distributions corresponding to different results can be seen in Figure 5.3. The α set $(1,1,1)$ corresponds to no numbers being generated yet and thus the Dirichlet shown a uniform distribution of which any probabilities are equally likely under the constraint that $p_1 + p_2 + p_3 = 1$. Whereas both $(5,5,5)$ and $50,50,50$ point towards a peak probability that the random number generator is fair, just as with the beta distribution, the distribution becomes sharper as we take more results.

You may have noticed that both the Dirichlet and beta distributions, especially in the cases of fair results, seem to mimic Gaussian distributions. This is not simply a visual coincidence but a direct consequence of Dirichlet's reduction to Gaussian's in the limit of large number counts and uncorrelated bins. Alternative approaches to describing the histogram nature of a $n(z)$ distribution that used a Gaussian approximation (see Appendix A in the work of Leistedt et al. (2016))

$$p(f_{ijk}|n_{ijk}) = \prod_{ijk} \mathcal{N}\left(\frac{n_{ijk}}{N_{\text{gal}}} \frac{n_{ijk}}{N_{\text{gal}}}\right). \tag{5.7}$$

This may seem like an appropriate approximation especially considering the visual evidence above. However, the different bins are inherently correlated through the $\sum_i p_i = 1$ constraint; if one probability goes up, the others must go down to compensate. Such lack of inter bin correlation can lead to further violations such as obtaining negative probability for our f_{ijk}'s; thus, this approximation is not usable compared to the full Dirichlet.

Other than being a conjugate prior and giving intuitive distributions that stay within the constraints, the other advantage of this model is the idea of 'Bayesian Shrinkage' or 'borrowing strength'. Putting the posterior for a single galaxy becomes more based on the population as a whole than on the individual galaxies' photometry. In this sense, one can deal with the issue of degeneracy as secondary peaks of individual galaxies become suppressed, especially if they indicate the galaxy exists in a low-population region. Consider

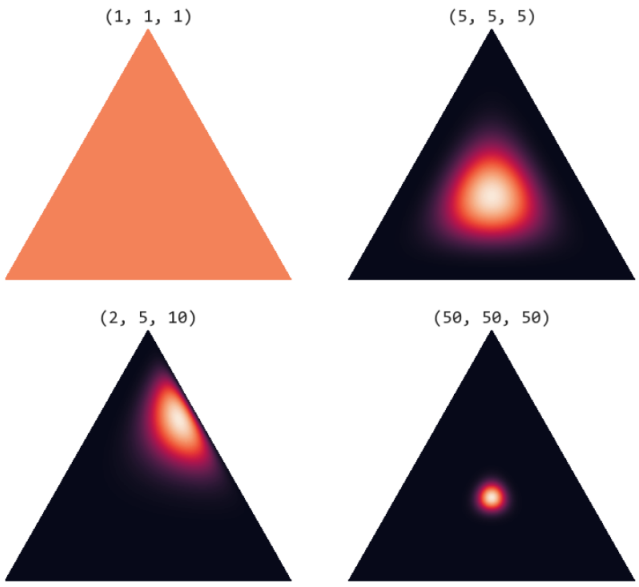


Figure 5.3: A set of Dirichlet distributions shown on the triangle of which the corners indicate a probability of $[1.0, 0.0, 0.0]$ for that outcome. The colour indicates the value of the probability on a scale of black to white, corresponding with 0.0 to 1.0 (Liu, 2022).

the differences in likelihood to posterior in the following z,t distributions inferred from the simulated data in the work of Leistedt et al. (2016) as seen in Figure 5.4. One can see that the strong degeneracy in the form of prominent secondary peaks in the likelihood function is suppressed by the posterior distribution due to the inferred population distribution. For more details of this suppression, please see Appendix B in the work of Leistedt et al. (2016).

While the above section presents the Bayesian Hierarchical model for inferring the redshift distributions, we now consider modification and extensions to consider critical ingredients needed for weak lensing analysis.

5.1.2 Template sensitivity

One of the fundamental assumptions of this model is that all galaxies within the survey are of the same template set used in the analysis. Thus, one of the critical issues is template incompleteness. One requires a large enough set of templates covering the photometric data's spectral range. This would point to using large enough template sets as possible. However, in past photometric codes, this can have detrimental effects because every template is as likely as any other. Take the following example if a set contained 6 elliptical templates and 1 spiral, you have, in a sense, introduced a prior favouring elliptical estimation. This imbalance can

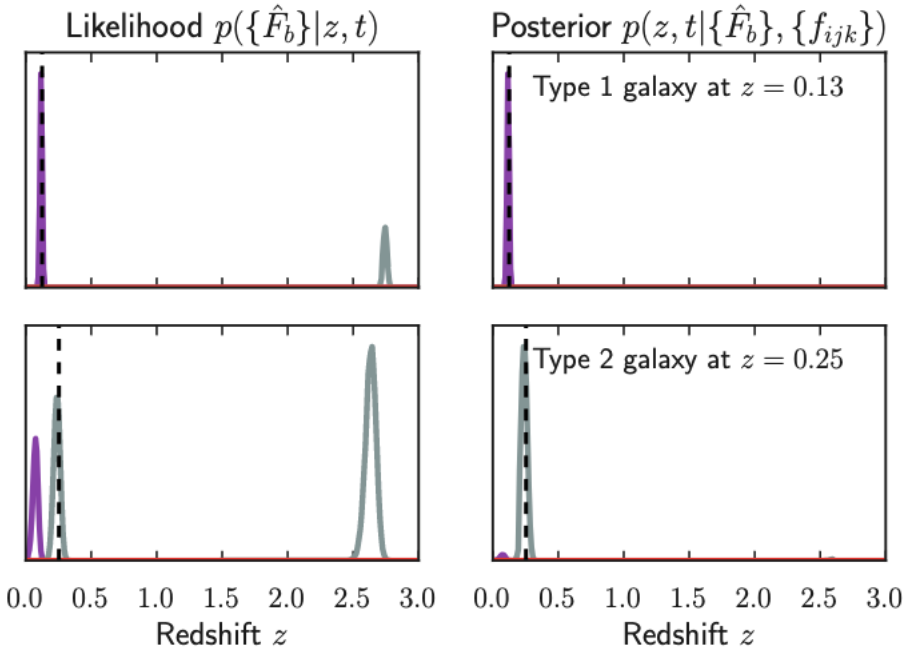


Figure 5.4: Likelihood and posterior distributions of 3 objects from the simulated set where the dashed line indicates the true redshift (Leistedt et al., 2016).

lead to catastrophic errors as different templates are susceptible to biases and degeneracies, allowing for differences in over-distribution results.

This issue does not hold for this method. As one can see from equation 5.5 the overall distribution is inferred from the cell number counts n_{ijk} of which a single galaxy can only be allocated to one cell (thus one template) at any one time. Similar templates (or even complete duplications) would be down-weighted due to the galaxy only being allocated to one of them at a time. In other words, if a single template represents the photometric survey, it will be up-weighted within the changing f_{ijk} regardless of the other surrounding templates.

With this in mind, the only issue restricting template size was the efficiency of which samples can be taken in future. However, this may not be an issue as much larger template sizes can be investigated. So instead, we looked for the smallest amount of templates that still allowed for the most prominent representation. This was achieved by using an aspect of the approach seen in the work of Ronen et al. (1999); taking a collection of 215 templates (a list of names of said templates can be seen in Table 5.1) used across photo-z tools including BPZ, Le-Phare (Ilbert et al., 2006), Hyper-Z (Bolzonella et al., 2000) and EAZY (Brammer et al., 2008) code and running them through principled component analysis (PCA).

PCA, first mentioned by Pearson (1901) and Hotelling (1933) is a technique to which one given a set of μ dimensional data creates a set of variables that are linear combinations of the original dimensions so that in this new variable space variance is maximised, and correlation

Table 5.1: Collection of 215 tables used in PCA analysis.

Template Name		
CWW_E_ext_Hyperz.sed	SB11_A_0_Calzetti_EBV0.5.sed	SB9_A_0_Calzetti_EBV0.3.sed
CWW_Im_ext_Hyperz.sed	SB11_A_0_nodust.sed	SB9_A_0_Calzetti_EBV0.4.sed
CWW_Sbc_ext_Hyperz.sed	SB1_A_0.sed	SB9_A_0_Calzetti_EBV0.5.sed
CWW_Scd_ext_Hyperz.sed	SB1_A_0_Calzetti_EBV0.1.sed	SB9_A_0_nodust.sed
El_cww.sed	SB1_A_0_Calzetti_EBV0.2.sed	Sb_A_0.sed
El_cww_DES_corr.sed	SB1_A_0_Calzetti_EBV0.3.sed	Sb_A_0_Calzetti_EBV0.1.sed
El_cww_capak.sed	SB1_A_0_Calzetti_EBV0.4.sed	Sb_A_0_Calzetti_EBV0.2.sed
Ell1_A_0.sed	SB1_A_0_Calzetti_EBV0.5.sed	Sb_A_0_Calzetti_EBV0.3.sed
Ell1_A_0_Calzetti_EBV0.1.sed	SB1_A_0_nodust.sed	Sb_A_0_Calzetti_EBV0.4.sed
Ell1_A_0_Calzetti_EBV0.2.sed	SB2_A_0.sed	Sb_A_0_Calzetti_EBV0.5.sed
Ell1_A_0_Calzetti_EBV0.3.sed	SB2_A_0_Calzetti_EBV0.1.sed	Sb_A_0_nodust.sed
Ell1_A_0_Calzetti_EBV0.4.sed	SB2_A_0_Calzetti_EBV0.2.sed	Sb_A_1.sed
Ell1_A_0_Calzetti_EBV0.5.sed	SB2_A_0_Calzetti_EBV0.3.sed	Sb_A_1_Calzetti_EBV0.1.sed
Ell2_A_0.sed	SB2_A_0_Calzetti_EBV0.4.sed	Sb_A_1_Calzetti_EBV0.2.sed
Ell2_A_0_Calzetti_EBV0.1.sed	SB2_A_0_Calzetti_EBV0.5.sed	Sb_A_1_Calzetti_EBV0.3.sed
Ell2_A_0_Calzetti_EBV0.2.sed	SB2_A_0_nodust.sed	Sb_A_1_Calzetti_EBV0.4.sed
Ell2_A_0_Calzetti_EBV0.3.sed	SB2_B2004a.norm.sed	Sb_A_1_Calzetti_EBV0.5.sed
Ell2_A_0_Calzetti_EBV0.4.sed	SB2_kin.sed	Sb_A_1_nodust.sed
Ell2_A_0_Calzetti_EBV0.5.sed	SB2_kin_DES_corr.sed	Sbc_cww.sed
Ell3_A_0.sed	SB2_kin_capak.sed	Sbc_cww_DES_corr.sed
Ell3_A_0_Calzetti_EBV0.1.sed	SB2_kin_mod3_Hyperz.sed	Sbc_cww_capak.sed
Ell3_A_0_Calzetti_EBV0.2.sed	SB3_A_0.sed	Sc_A_0.sed
Ell3_A_0_Calzetti_EBV0.3.sed	SB3_A_0_Calzetti_EBV0.1.sed	Sc_A_0_Calzetti_EBV0.1.sed
Ell3_A_0_Calzetti_EBV0.4.sed	SB3_A_0_Calzetti_EBV0.2.sed	Sc_A_0_Calzetti_EBV0.2.sed
Ell3_A_0_Calzetti_EBV0.5.sed	SB3_A_0_Calzetti_EBV0.3.sed	Sc_A_0_Calzetti_EBV0.3.sed
Ell4_A_0.sed	SB3_A_0_Calzetti_EBV0.4.sed	Sc_A_0_Calzetti_EBV0.4.sed
Ell4_A_0_Calzetti_EBV0.1.sed	SB3_A_0_Calzetti_EBV0.5.sed	Sc_A_0_Calzetti_EBV0.5.sed
Ell4_A_0_Calzetti_EBV0.2.sed	SB3_A_0_nodust.sed	Sc_A_0_nodust.sed
Ell4_A_0_Calzetti_EBV0.3.sed	SB3_B2004a.norm.sed	Sc_A_1.sed
Ell4_A_0_Calzetti_EBV0.4.sed	SB3_kin.sed	Sc_A_1_Calzetti_EBV0.1.sed
Ell4_A_0_Calzetti_EBV0.5.sed	SB3_kin_DES_corr.sed	Sc_A_1_Calzetti_EBV0.2.sed
Ell5_A_0.sed	SB3_kin_capak.sed	Sc_A_1_Calzetti_EBV0.3.sed
Ell5_A_0_Calzetti_EBV0.1.sed	SB3_kin_mod3_Hyperz.sed	Sc_A_1_Calzetti_EBV0.4.sed
Ell5_A_0_Calzetti_EBV0.2.sed	SB4_A_0.sed	Sc_A_1_Calzetti_EBV0.5.sed
Ell5_A_0_Calzetti_EBV0.3.sed	SB4_A_0_Calzetti_EBV0.1.sed	Sc_A_1_nodust.sed
Ell5_A_0_Calzetti_EBV0.4.sed	SB4_A_0_Calzetti_EBV0.2.sed	Sc_A_2.sed
Ell5_A_0_Calzetti_EBV0.5.sed	SB4_A_0_Calzetti_EBV0.3.sed	Sc_A_2_Calzetti_EBV0.1.sed
Ell6_A_0.sed	SB4_A_0_Calzetti_EBV0.4.sed	Sc_A_2_Calzetti_EBV0.2.sed
Ell6_A_0_Calzetti_EBV0.1.sed	SB4_A_0_Calzetti_EBV0.5.sed	Sc_A_2_Calzetti_EBV0.3.sed
Ell6_A_0_Calzetti_EBV0.2.sed	SB4_A_0_nodust.sed	Sc_A_2_Calzetti_EBV0.4.sed
Ell6_A_0_Calzetti_EBV0.3.sed	SB5_A_0.sed	Sc_A_2_Calzetti_EBV0.5.sed
Ell6_A_0_Calzetti_EBV0.4.sed	SB5_A_0_Calzetti_EBV0.1.sed	Sc_A_2_nodust.sed
Ell6_A_0_Calzetti_EBV0.5.sed	SB5_A_0_Calzetti_EBV0.2.sed	Scd_cww.sed
Ell7_A_0.sed	SB5_A_0_Calzetti_EBV0.3.sed	Scd_cww_DES_corr.sed
Ell7_A_0_Calzetti_EBV0.1.sed	SB5_A_0_Calzetti_EBV0.4.sed	Scd_cww_capak.sed
Ell7_A_0_Calzetti_EBV0.2.sed	SB5_A_0_Calzetti_EBV0.5.sed	Sd_A_0.sed
Ell7_A_0_Calzetti_EBV0.3.sed	SB5_A_0_nodust.sed	Sd_A_0_Calzetti_EBV0.1.sed
Ell7_A_0_Calzetti_EBV0.4.sed	SB6_A_0.sed	Sd_A_0_Calzetti_EBV0.2.sed
Ell7_A_0_Calzetti_EBV0.5.sed	SB6_A_0_Calzetti_EBV0.1.sed	Sd_A_0_Calzetti_EBV0.3.sed
Im_cww.sed	SB6_A_0_Calzetti_EBV0.2.sed	Sd_A_0_Calzetti_EBV0.4.sed
Im_cww_DES_corr.sed	SB6_A_0_Calzetti_EBV0.3.sed	Sd_A_0_Calzetti_EBV0.5.sed
Im_cww_capak.sed	SB6_A_0_Calzetti_EBV0.4.sed	Sd_A_0_nodust.sed
SB0_A_0.sed	SB6_A_0_Calzetti_EBV0.5.sed	Sd_A_1.sed
SB0_A_0_Calzetti_EBV0.1.sed	SB6_A_0_nodust.sed	Sd_A_1_Calzetti_EBV0.1.sed
SB0_A_0_Calzetti_EBV0.2.sed	SB7_A_0.sed	Sd_A_1_Calzetti_EBV0.2.sed
SB0_A_0_Calzetti_EBV0.3.sed	SB7_A_0_Calzetti_EBV0.1.sed	Sd_A_1_Calzetti_EBV0.3.sed
SB0_A_0_Calzetti_EBV0.4.sed	SB7_A_0_Calzetti_EBV0.2.sed	Sd_A_1_Calzetti_EBV0.4.sed
SB0_A_0_Calzetti_EBV0.5.sed	SB7_A_0_Calzetti_EBV0.3.sed	Sd_A_1_Calzetti_EBV0.5.sed
SB0_A_0_nodust.sed	SB7_A_0_Calzetti_EBV0.4.sed	Sd_A_1_nodust.sed
SB10_A_0.sed	SB7_A_0_Calzetti_EBV0.5.sed	Sd_A_2.sed
SB10_A_0_Calzetti_EBV0.1.sed	SB7_A_0_nodust.sed	Sd_A_2_Calzetti_EBV0.1.sed
SB10_A_0_Calzetti_EBV0.2.sed	SB8_A_0.sed	Sd_A_2_Calzetti_EBV0.2.sed
SB10_A_0_Calzetti_EBV0.3.sed	SB8_A_0_Calzetti_EBV0.1.sed	Sd_A_2_Calzetti_EBV0.3.sed
SB10_A_0_Calzetti_EBV0.4.sed	SB8_A_0_Calzetti_EBV0.2.sed	Sd_A_2_Calzetti_EBV0.4.sed
SB10_A_0_Calzetti_EBV0.5.sed	SB8_A_0_Calzetti_EBV0.3.sed	Sd_A_2_Calzetti_EBV0.5.sed
SB10_A_0_nodust.sed	SB8_A_0_Calzetti_EBV0.4.sed	Sd_A_2_nodust.sed
SB11_A_0.sed	SB8_A_0_Calzetti_EBV0.5.sed	Sdm_A_0.sed
SB11_A_0_Calzetti_EBV0.1.sed	SB8_A_0_nodust.sed	Sdm_A_0_Calzetti_EBV0.1.sed
SB11_A_0_Calzetti_EBV0.2.sed	SB9_A_0.sed	Sdm_A_0_Calzetti_EBV0.2.sed
SB11_A_0_Calzetti_EBV0.3.sed	SB9_A_0_Calzetti_EBV0.1.sed	Sdm_A_0_Calzetti_EBV0.3.sed
SB11_A_0_Calzetti_EBV0.4.sed	SB9_A_0_Calzetti_EBV0.2.sed	Sdm_A_0_Calzetti_EBV0.4.sed
SB11_A_0_Calzetti_EBV0.5.sed	SB9_A_0_Calzetti_EBV0.3.sed	Sdm_A_0_Calzetti_EBV0.5.sed
SB11_A_0_nodust.sed	SB9_A_0_Calzetti_EBV0.4.sed	Sdm_A_0_nodust.sed

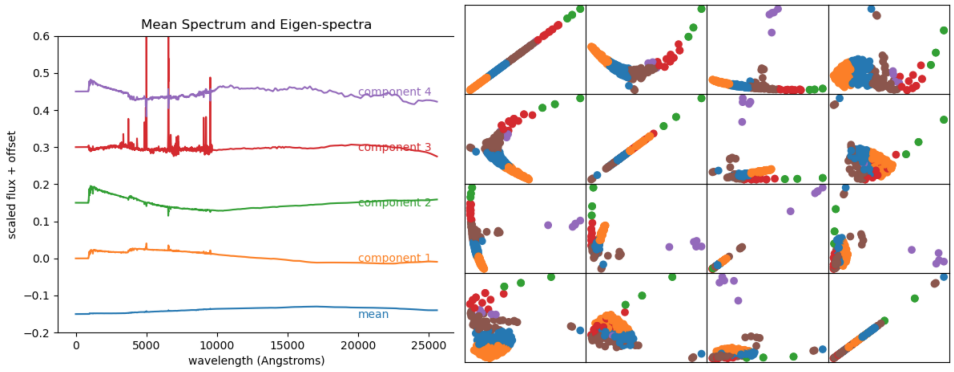


Figure 5.5: *left*: The mean of the first 4 components extracted from PCA from a large template set, the flux has been scaled and offset for visual purposes. *right*: A 4x4 scatter plot of all templates plotted against component i,j (where i=[1,4]). The colours indicate the different sets of which the K mean clustering algorithm placed the template.

minimised. This is achieved by considering a data $n \times \mu$ data set $\mathbf{X}$ of which n is the number of individual data points. The value of a new variable based on a linear combination of the original variables would be given as $\mathbf{Xa}$ where $\mathbf{a} = [a_1, ..., a_\mu]$ is a set of coefficients and the variance of this new variable by $var(\mathbf{Xa}) = \mathbf{a}^T\mathbf{Sa}$ where $\mathbf{S}$ is the covariance matrix of the original data.We are then interested in maximising $\mathbf{a}^T\mathbf{Sa}$ which when combined with the constraint $\mathbf{a}^T\mathbf{a} = 1$ (unit-norm vectors) can be transformed into an eigenvalues problem (Jollife and Cadima, 2016)

$$\mathbf{Sa} = \lambda\mathbf{a}, \tag{5.8}$$

where the coefficients are the eigenvectors and λ the variance as $\mathbf{a}^T\mathbf{S} = \lambda\mathbf{a}^T\mathbf{a} = \lambda$, the eigenvalues can then be ranked in order of size in order to keep as many principled components as required.

Using the sklearn PCA package, we uncovered 4 components to which a linear combination of the 4 best describes the complete template set. The 4 components can be seen in Figure 5.5 *left* to which prominent characteristics of specific templates such as emission lines can be visible. The reason for this step was to use PCA as a form of data reduction as it has been shown that machine learning algorithms based around classification train quicker and perform more accurately when using PCA data as opposed to raw data, especially when the raw data is spectra (Howley et al., 2006). Then, a k clustering algorithm was used to group the templates into 5 sets. The algorithm looks to 5 points on PCA variable space to which the objects classified to a certain point are closer than any other data points. Figure 5.5 shows that the sets are pretty distinctively grouped based on the PCA components. From these groups, an equal amount of 10 templates were randomly sampled from the individual groups

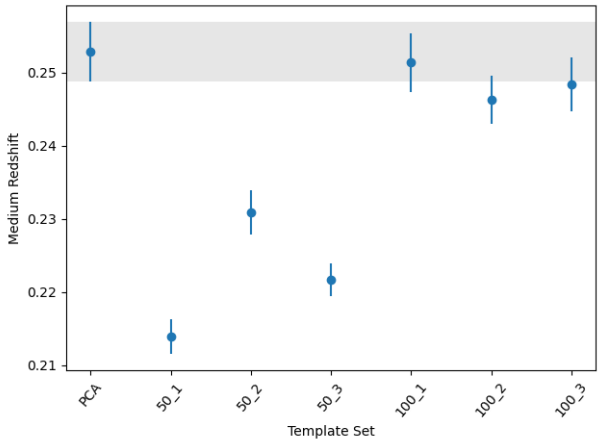


Figure 5.6: The median redshift based on the redshift distribution of the KiDS450 -Viking data in the first tomographic bin and their variance compared to 6 random template sets (3 containing 50 templates and 3 containing 100 templates), these sets are compared to the same result gathered from PCA analysis (grey). This is the same results as displayed in the first tomographic bin of 6.7 *left*

giving a set of 50 templates. To test how well this set represents the photometric survey, the redshift distribution analysis will be taken with different sets of randomly selected templates to examine the level of variability as well as random sets of much larger template sizes, ones of which minimal variability is expected due to there completeness but are impractical to use on large scales. By testing only the first tomographic bin as seen in Figure 5.6 the use of the 50 PCA templates seems promising as the medium redshift is in agreement with the random 100 template sets while the random 50 template sets show to much variance in there results to be of use. For further validation the the same test will take place across all tomographic bins in the final result.

5.1.3 Selection Effects

The model assumes that outside of the reference band magnitude (a dimension of the distribution p(t,z,m)), no selection effects are being applied to the galaxy sample and that the survey is a random sample of the actual distribution. Not accounting for these selection effects can lead to artificial peaks in the redshift distribution. To understand our method for dealing with the selection effects of a typical weak lensing survey, it is first worth exploring the approach suggested by the original paper and why it would not be appropriate for our needs.

As stated by Leistedt et al. (2016) the accounting of said effects can be applied by modifying the likelihood to include a selection term s for whether a galaxy is included in the survey.

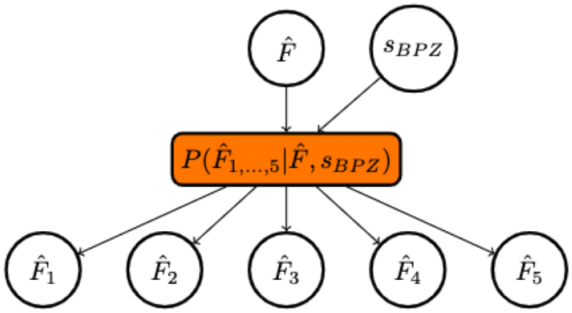


Figure 5.7: Hierarchical forward model for the selection effects of all possible observed fluxes $\hat{F}$ to those kept within the survey in their separate tomographic bins $\hat{F}_{1...5}$. The selection effect s_{BPZ} is the categorising of sources by BPZ redshift estimates z_{BPZ}. The b superscripts have been omitted.

$$p(\{\hat{F}_b\}|s,t,z,m,\sigma) = \frac{p(\{\hat{F}_b\}|t,z,m,\sigma)p(s|\{\hat{F}_b\})}{p(s|t,z,m,\sigma)}. \tag{5.9}$$

In most cases $p(s|\{\hat{F}_b\})$ is binary and ignored leaving the inclusion of selection with the denominator

$$p(s|t,z,m,\sigma) = \int p(s|\{\hat{F}_b\})p(\{\hat{F}_b\}|t,z,m,\sigma)d\hat{F}_1...d\hat{F}_{N_b}. \tag{5.10}$$

The most prominent selection to take place in weak lensing surveys comes through the categorising of data into tomographic bins. This is achieved by finding their photo-z estimate $\hat{z}_{\rm BPZ}(\{\hat{F}_b\})$ (most probable Bayesian redshift from BPZ) and matching the estimates to their appropriate bin. An effect of this kind can be demonstrated as an extension to Figure 5.1 through the selection variable s as shown in Figure 5.7. Such a selection would have a complex function such as

$$p(s_i|\{\hat{F}_b\}) = \Theta(\hat{z}_{\rm BPZ}(\{\hat{F}_b\}) - z_{{\rm min}_i})\Theta(z_{{\rm max}_i} - \hat{z}_{\rm BPZ}(\{\hat{F}_b\})). \tag{5.11}$$

Due to the photo-z estimate, $\hat{z}_{BPZ}(\{\hat{F}_b\})$ is based on a code. There is no analytic solution for $p(s|z,t,m)$ (as opposed to a selection based on flux limit where the analytical solution $p(s|t,z,m,\sigma) = 2^{-N_b}\prod_b^{N_b}(1-erf((F_{{\rm min}_b} - F_{{\rm mod}_b}(t,z,m))/\sigma\sqrt{2}))$ exists). Instead, an integration of equations 5.9 and 5.10 can, through the Monte-Carlo approach which one simulates fluxes at each (i,j,k) point of the histogram and observe how many are attached to each tomographic bin using said BPZ criteria.

The approach described above is what was used in the original paper. However, in weak lensing surveys, one answers a slightly different question than the above. We are not looking

for the underlying distribution of which our survey is a sample of ie.f_{ijk} instead. We are looking for the distribution *of that survey* which is provided to us through the parameter n_{ijk}. The subtle change in the parameter substantially affects what one records as the n_{ijk} is not as directly influenced by the uniform distribution prior as f_{ijk}. One may argue that this is not an issue at high populations seen in weak lensing surveys as the posteriors would not be prior dominated, and the distribution would approximate to a n_{ijk} dictated Gaussian as seen in equation 5.7 but this is also not the case. A minor reason would be that even if the entire population is high, if the population per t,z,m cell is low, it would still be influenced by the uniform prior in those cells and produce inaccuracies even in regions of high S/N.

A major reason is that this change in parameter also changes our use of selection effects. When looking for the underlying distribution, one must mitigate all selection effects as seen in equation 5.7 whereas when we look for the distribution of the survey itself, any selection effect should also be included. For clarity consider Figures 5.1 and 5.7 combined. If all the selection effects were correctly accounted for, no matter which tomographic bin of data $\hat{F}_i$ one was analysing, one would still end up with the same distribution f_{ijk} be it with different levels of uncertainty.

That said, one wants to maximise the available information, so one does not apply the method to each tomographic bin individually, as the sources provided in equation 2 are both a smaller number and heavily skewed towards their selection criteria. Instead, the method is applied to the entire population to get the distribution of the entire survey, and then the selection effects are applied afterwards to get the individual populations.

This can be achieved in one of 2 ways, if one knows the probability of the source of a particular z,t,m being selected into one of the tomographic bins s_i, i.e. $p(s_i|z,t,m)$ One can sample from a multinomial distribution where the mean distribution of a tomographic bin is given as

$$\bar{n}_{ijk,\text{bin}} = n_{ijk,\text{survey}} \times p(s_{\text{bin}}|z,t,m), \tag{5.12}$$

however, this approach has weaknesses:

1. As explained above, there is no analytical solution for $p(s_i|z,t,m)$, so one relies on Monte-Carlo simulations to replicate the distribution. Although such a simulation would require every detail between a source being observed and placed into a bin to be effective as the simulation needs to replicate the pipeline, this kind of information might not always be available, let alone public.

2. For the Monte Carlo simulation to be effective, it would also require the sources included within the simulation to replicate the sources seen within the survey, both in

fluxes and flex errors. Such a simulation would be doubling down on the assumption that every source within the survey can be described entirely through the z, t, m as one will be simulating sources for z, t, m cells of which the actual survey might be of the low population the simulated sources may not be a good representation of the sources in that region.

3. Alternatively, as long as one has access to the selected sources, one can apply their selections directly. This can be achieved by no longer using n_{ijk}. As stated, n_{ijk} is the number of sources allocated to a particular cell (in any particular sample). This summarises the actual parameter, allocating each source in each cell.

With the above in mind, the actual sample parameters of interest are the $\{i, j, k\}_g$ the magnitude, redshift and type allocation of each source from the first part of the Gibbs sampler in equation 5.4. By collecting these samples, one not only gains individual posterior distributions for each source by separating the sources by tomographic bin allocation and weighting them by their lens-fit weights; one can also gain samples for the individual redshift distributions.

One of the downsides of this approach is that one no longer gets the speed and memory benefits of only wanting the f_{ijk}'s; instead, every sample consists of $N_{\rm gal}$ data points leading to a heavy slow down in memory capacity and computational speed. In order to tackle these issues, choices had to be made that, on the surface, would lower the accuracy. The first was the cut in the number of templates used, this may increase template incompleteness, but due to our PCA method as described in 5.1.2, this can be mitigated. The second was to limit the number of galaxies used in the analysis to $100,000$. To be clear, this is only in the photometric portion of the analysis, not the cosmological portion.

Chapter 6

KV450 Analysis

With the hierarchical model set so that we can apply its use to topographical bins and the choice of templates selected, we are now in a position to apply our approach to a weak lensing survey. Both real and simulated.

6.1 Data

6.1.1 Simulation

To demonstrate the validity of the method and the new implementations beyond the original paper and in the realm of the KV-450 survey and set if simulated data was created.

To begin, $N_{\rm gal} = 100,000$ r-band magnitudes are drawn from a distribution replicating the r-band distribution from the KV-450 survey. Although the number is fewer factors than the survey, it is sufficient to replicate the issues with the high populations mentioned above. Secondly, low S/N sources are primarily associated with high magnitude sources, so replicating the actual survey's magnitude distribution is highly important to test the robustness. The allocation of true z,t to each true m was decided to test potential weaknesses in the method while staying realistic. A chosen amount of correlations were removed to minimise information on which populations can be formed in the Gibbs sampler. For example, types were drawn independently through a uniform distribution on t with the only changes being due to shot noise. The templates used were the same as mentioned above in Table 5.1. This was chosen to replicate the variability of real-world galaxy spectra and test the level of representation of our PCA template set. At the same time, actual z values were drawn from the following distribution:

$$p(z|m_i) = \frac{4}{5}\mathcal{N}(\mu_{1,i}, \sigma^2_{1,i}) + \frac{1}{5}\mathcal{N}(\mu_{2,i}, \sigma^2_{2,i}), \tag{6.1}$$

where μ, σ are the means and variance of the Gaussian distributions chosen to increase with magnitude bin to replicate some of the behaviour seen in the BPZ priors. The inclusion of the secondary minor Gaussian $\mathcal{N}(\mu_{2,i}, \sigma^2_{2,i})$ is necessary to test to what extent secondary peaks are suppressed even when they reflect reality, especially in regions of low S/N. Lastly, 5% of the simulated population was given redshifts independently of magnitude and type following a uniform distribution to test the method's ability to correctly allocate sources even when their values go against the major distributions of the population.

Based on the allocated true z,t,m true fluxes in the 9 filters (F_b where b is the filter) associated with KV-450 are drawn directly from the model

$$F_b(t,z) = \frac{1+z}{4\pi D_L^2(z)} \int_0^\infty \frac{d\nu}{nu} L_{\nu,t}[\nu(1+z)] W_b(\nu), \tag{6.2}$$

where $L_{\nu,t}(\nu)$ is the rest-frame luminosity density (as used by Leistedt et al. (2016)) of the extragalactic source of type t as a function of frequency ν and $W_b(\nu)$ are functions characterising the response of the photometric filters. Dividing by the r-band reveals the flux ratios, which are then multiplied by the r-band fluxes drawn above.

In order to simulate observable fluxes $\hat{F}_b$ are drawn from a Gaussian distribution $\mathcal{N}(F_{b,g}, \sigma^2_{b,g})$. Where $\sigma_{b,g}$ is the variances on a flux given its filter b and source g. To best represent errors seen in actual KV-450 data, a neural network containing 100 hidden layers through sklearn.neural_network.MLPRegressor was trained on KV-450 data. Where the inputs and outputs were the observed fluxes and their errors for each source, respectively. A similar approach was used for applying lens-fit weights to each source (though this network could be improved through the inclusion of size data on the sources). A similar neural network shown from the one above was used. In this case, the inputs were the fluxes and the errors of each source, while the outputs were the lens-fits weights trained on the same data above.

Allocating each source to the correct tomographic bin took a more direct approach. As stated previously, a source is allocated to a tomographic bin depending on its photo-z estimation $\hat{z}_{BPZ}$ which is the mean redshift from the BPZ redshift posteriors. However, the simulated observed fluxes and their errors could not be inputted directly into BPZ but instead followed the same pipeline of KV-450. Sources in the KV-450 pipeline were applied to BPZ as magnitudes, not fluxes so the data had to be converted using $m_b = -2.5\log_{10}(\hat{F}_b)$ for the flux to magnitude conversion and $m_{e,b} = 1.086 \times \sigma_b/\hat{F}_b$ to convert flux errors to magnitude errors to a first order approximation. Sources have filters classed as 'unobserved' if they fall under 1 of 2 conditions, ether $\sigma_b > \hat{F}_b$ or $m > lim_b$ where lim_b is the magnitude limit for a certain panel of the sky, these limits were randomly allocated to sources. Although technically these sources are 'undetected', due to a bug in the original survey they were

classed as unobserved (in reality information can still be gathered from undetected sources due to flux limits), the simulation looks to replicate the BPZ data of the original survey and thus the bug is also replicated here. The classification of 'unobserved' means setting $m = 99$ BPZ and then treating the filter to have the maximum error, and no information can be gathered from it. BPZ also takes a M_0 input. KV-450 data is labelled 'MAG_AUTO' and consists of a non-linear combination of r and i filters and their errors and extractions. It is the M_0 value that interacts with the BPZ priors. M_0 values were also simulated using a trained neural network where $m_{r,i}$ are the inputs and M_0 the output.

BPZ mean distributions $\hat{z}_{BPZ}$ aloud the allocation of each source to each tomographic bin based on the following categories $\hat{z}_{BPZ} = [0.1, 0.3], [0.3, 0.5], [0.5, 0.7], [0.7, 0.9], [0.9, 1.2]$. To justify the method, the simulated survey is analysed using the same template set to which the survey was created.

6.1.2 KiDS + VIKING 450

The data used for the cosmological parameter inference and demonstration of the application of this Bayesian hierarchical model was the KiDS+VIKING-450 data set (Kannawadi et al., 2019; Wright et al., 2019). This combines optical (KiDS) and near-infrared (VIKING) photometric data. KiDS is a wide-angle photometric survey which uses the OmegaCAM camera within the VLT Survey Telescope (de Jong et al., 2012). With a target area of $\sim$ 1500 deg^2, the survey intends to reveal cosmological information through the large-scale distribution of matter. Although multiple data realises have taken place (de Jong et al., 2015; De Jong et al., 2017) KiDS-450 refers to DR3 of which 450 deg^2 of sources in the *ugri* bands have become available.

Infrared data is collected from the VISTA Kilo-degree infrared Galaxy survey, which used the Visible and Infrared CAMera (VIRCAM) on ESO's 4m VISTA telescope (Venemans et al., 2015). The planned target is for VIKING to cover $\sim$ 1350 deg^2 of sky across the $ZYJHK_s$ IFR bands. However, mainly due to masking and slightly due to incompleteness in the coverage of VIKING, to ensure that the combined data set of KV-450 is fully covered in all nine bands, it has a reduced unmasked area to 341 deg^2.

The specific data used was the Gaussian Aperture and PSF magnitudes (GAAP), a post-processing stage of data reduction unchanged from previous releases (Kuijken, 2008; Kuijken et al., 2015). However, other forms of multi-band data are available, including the GAAP magnitude/fluxes and the default used by KiDS. As the Bayesian Hierarchical Model (BHM) method is much more heavily reliant on photometric data for $n(z)$ estimation than that of DIR (which also has spectroscopic data), another data filtering process took place to ensure optimal photo-z estimations. The process replicates that of other recent photo-z studies

(Bilicki et al., 2018). Sources removed that do not include magnitude errors in every band, as well as removing any source that is flagged by requiring GAAP_Flag_ugriZYJHKs==0. This requirement reduced the KV-450 area by $\sim 5\%$. We have not used the GAAP magnitudes, unlike the analysis of KiDS for their DIR and BPZ pipelines for full redshift distribution and tomographic bin location. Instead, the GAAP fluxes have been used as our input. These are available in the KiDS data set as 'FLUX_GAAP_b' and 'FLUXERR_GAAP_b' for the fluxes and their errors, respectively, where 'b' is the 9 filters. The fluxes need to be corrected before analysis using offset and extinction terms

$$F_{\text{corrected},b} = F_b \times 10^{\frac{2}{5}(\text{extinction}_b - \text{offset}_b)}, \tag{6.3}$$

where $F_{\text{corrected},b}$ is the final flux (although equation 6.3 applies the same to the flux errors) used for analysis and F_b the catalogued flux. The extinction term is taken directly from the 'EXTINCTION_b' column of the data set. The offset term exists as we do not have access to all the corrections that need to be calculated. T approximately recover those said corrections sources are grouped by the tile of the sky in which they were observed. This is found under 'THELI_NAME'. Sources within the same tile have the same offset, calculated as the median value of the following quantity

$$\text{offset}_b = \text{MAG\_GAAP\_b} + 2.5\log_{10}(\text{FLUX\_GAAP\_b}). \tag{6.4}$$

Using the median is particularly important as one can not calculate the individual offsets for sources that have been classified as 'undetected' or 'unobserved'. The tomographic bin allocation is determined by the BPZ values reported in the original data' Z_b', whereas the weights are also taken directly from 'recal_weight'. The galaxies are not premade from particular templates as this is actual data. With this in mind, the entire catalogue of over 300 templates available through BPZ was used for the analysis.

6.2 Cosmological Analysis

This section expands on the data products used and the forward theoretical modelling. In short, we replicate the KV-450 analysis with a new set of redshift distributions, as inferred using our method. All other data products, such as the data vector and the covariance matrix, are similar to the ones used by Hildebrandt et al. (2020). See Figures 6.2 and 6.1 for reference. We will discuss the data in §6.2.2, the forward modelling approach in §6.2.3 before touching briefly on the different systematics in §6.2.4. We refer the reader to the work of Hildebrandt et al. (2020) for a more detailed and comprehensive study.

6.2.1 Sample Implementation

Even after $n(z)$ samples have been correctly taken from the Gibbs sampling, how one implements them into the cosmological analysis also has to be chosen. There are multiple ways in which the $n(z)$ samples could have been implemented into the parameter inference. We consider two different methods, both of which have been implemented in previous KiDS studies; they have advantages and disadvantages. Firstly, the method is similar to the ones used in the original KV-450 analysis (Hildebrandt et al., 2018), where the mean redshift and error on this means are extracted from the $n(z)$ samples for each tomographic bin. This mean and error are then used on the μ and σ in a Gaussian prior for a set of nuisance parameters δz_i (where i is the tomographic bin). The nuisance parameter defines a linear shift in z for the used $n_i(z)$ in the parameter inference

$$n_i(z) \rightarrow n_i(z + \delta z_i). \tag{6.5}$$

Alternatively, one can use the samples directly. See, for example, the KiDS-450 analysis (Hildebrandt et al., 2017). In this case, a different sample of $n_i(z)$ is used with every likelihood evaluation. The linear method contains the nuisance parameters allowing the $n(z)$ to converge on the correct distribution. On the other hand, sampling directly has its benefits. Firstly, the explored parameter space is smaller by not including the nuisance parameters, leading to more efficient inference. Secondly, the $n(z)$ samples allow for inter-bin correlations that the linear shift would ignore. This usually forms significant anti-correlation between adjacent inter-bins, an expected property of the Dirichlet model. Thirdly, the convergence issue would not apply to the BHM model as the chain of $n(z)$ samples have already converged to a correct distribution for the $n(z)$ as shown through the Gelman-Rubin test. We also consider merging the two methods and applying them to the original KV-450 analysis to see if the results remain consistent. Finally, we use the linear shift method to create 1000 samples of the $n(z)$ gaining our δz_i values straight from the Gaussian prior, then use the samples directly as with the case of the second method. The advantage is that we get the more generous error associated with the linear shift but the efficiency of having fewer parameters associated with the direct method. All three methods (random samples, nuisance parameters, hybrid) were compared through parameter inference, and no statistically significant difference is visible. However, using the samples directly means one is not forcing particular distributions onto their $n(z)$ uncertainties but instead taking on the uncertainties as they appear to form the data, which may affect future works in which data quality varies at different redshifts and magnitudes, with no computational efficiency lost it was decided that the raw samples were the chosen implementation route.

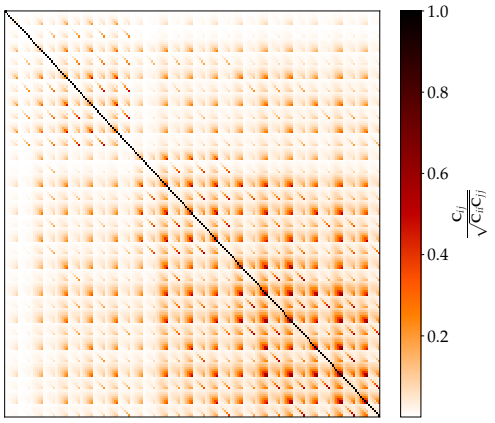


Figure 6.1: The data correlation matrix used in this analysis. Each element in the original covariance matrix, $C \in \mathscr{R}^{195\times195}$ is scaled such that the maximum possible entry is 1 for illustration.

6.2.2 Data

The original KV-450 analysis employs the TREECORR code to estimate the shear correlation function between two tomographic bins i and j, that is,

$$\hat{\xi}^{ij}_{\pm}(\theta) = \frac{\sum_{ab} w_a w_b \left[\varepsilon^i_t(x_a)\varepsilon^j_t(x_b) \pm \varepsilon^i_\times(x_a)\varepsilon^j_\times(x_b)\right]}{\sum_{ab} w_a w_b} \tag{6.6}$$

where w is the *lensfit* weight, ε_t and $\varepsilon_\times$ correspond to the tangential and cross ellipticities of a galaxy with reference to the vector $x_a - x_b$ between a pair of galaxies a and b. Nine logarithmically spaced bins are used within the interval $\theta \in [0'.5, 300']$. We follow the same procedure as used by Hildebrandt et al. (2020) and use the first seven bins for ξ_+ and the last six bins for ξ_-. Since we have five tomographic redshift distributions, we have five auto and ten cross pairs to tomographic redshift bins. The total number of data points is $15 \times (7+6) = 195$. In Figure 6.2, we show the 195 data points and their associated uncertainties. Hildebrandt et al. (2020) also generated the data covariance matrix via an analytical recipe, and this covariance matrix, normalised such that the maximum entry is one, is shown in Figure 6.1. In §6.2.3, we cover the forward theoretical model, and throughout this work, we assume a Gaussian likelihood for the data with a fixed covariance matrix, that is,

$$-2\log L = (d-\mu)^T C^{-1}(d-\mu) + \log|C| + \text{constant}. \tag{6.7}$$

6.2.3 Theory

The two-point correlation functions between bins i and j are used as the forward cosmological model to model the data. It is strictly a linear combination of the E- and B-mode power spectra, that is,

$$\xi_{\pm}^{ij}(\theta) = \int_0^{\infty} d\ell \frac{\ell}{2\pi} J_{0/4}(\ell\theta)[C_{\ell,ij}^{EE} C_{\ell,ij}^{BB}] \tag{6.8}$$

and the Bessel functions of the first kind, $J_{0/4}$ are the weight functions. In the absence of systematics, the B-mode power spectra are assumed to be zero, and the main contribution is due to the convergence power spectra, $C_{\ell,ij}^{EE} = C_{\ell,ij}^{\kappa\kappa}$. Under the Limber approximation (Loverde and Afshordi, 2008) and assuming a flat universe. The EE power spectrum is given by

$$C_{\ell,ij}^{\mathrm{EE}} = \int_0^{\chi_{\mathrm{H}}} \mathrm{d}\chi \, \frac{w_i(\chi)\, w_j(\chi)}{\chi^2} P_{\delta}(k;\chi) \tag{6.9}$$

where $P_{\delta}(k;\chi)$ is the non-linear matter power spectrum. The weight function $w(\chi)$ is

$$w_i(\chi) = A\chi(1+z) \int_{\chi}^{\chi_{\mathrm{H}}} \mathrm{d}\chi' \, n_i(\chi) \left(\frac{\chi' - \chi}{\chi'} \right) \tag{6.10}$$

and $A = 3H_0^2\Omega_m/2c^2$. The weight functions, w depends on the redshift distribution, $n_i(z)\,dz = n_i(\chi)\,d\chi$. The redshift distributions are also properly normalised such that

$$\int n(z)\,dz = 1.$$

In this analysis, we sample five cosmological parameters, which we denote as Φ:

$$\Phi = \left[\Omega_{cdm}h^2, \ln 10^{10} A_s, \Omega_b h^2, n_s, h\right]$$

and two massless neutrinos and one massive neutrino are assumed. The latter is fixed at $m_\nu = 0.06\,eV$. We also record two derived quantities, namely Ω_m and $S_8 \equiv \sigma_8\sqrt{\Omega_m/0.3}$. Sampling is done in `MontePython` (Brinckmann and Lesgourgues, 2018), which also wraps `CLASS` (Blas et al., 2011). Apart from the cosmological parameters, we also have to account for nuisance parameters, which we discuss briefly in the next section.

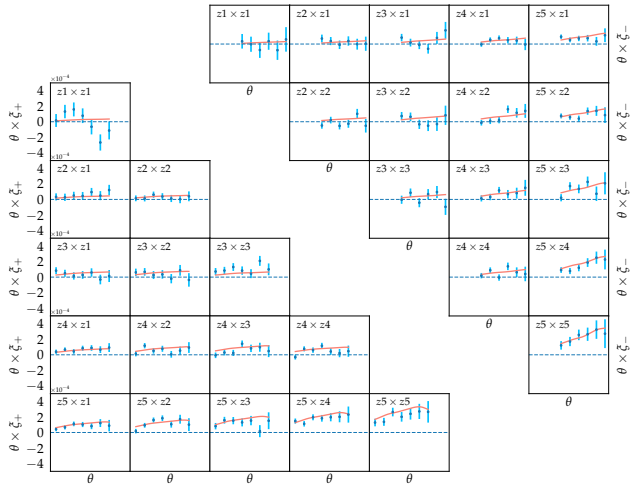


Figure 6.2: The correlation functions multiplied by the angular scales, θ. We have seven ξ_+ and six ξ_- per pair of tomographic redshift distributions, which results in a total of $13 \times 15 = 195$ correlation functions. Hence, the data vector, $x \in \mathscr{R}^{195}$ is shown in blue above. The error bars correspond to the standard deviation, computed using the diagonal elements of the covariance matrix. The red curve shows the fit to the data using the mean of the inferred cosmological and nuisance parameters with the Combined set.

6.2.4 Systematics

The shear lensing signal is a biased tracer of the convergence power spectrum. The model which is commonly used in recent weak lensing analyses incorporates additional terms due to intrinsic alignments, II and GI, that is,

$$C^{\rm tot}_{\ell,ij} = C^{\kappa\kappa}_{\ell,ij} + A^2_{IA} C^{II}_{\ell,ij} - A_{IA} C^{GI}_{\ell,ij}. \tag{6.11}$$

The II term, which arises due to the correlation of ellipticities in the local environment, contributes positively to the total lensing signal. In contrast, the GI term arises due to the correlation between intrinsic ellipticities of foreground and gravitational shear of background galaxies (Hirata and Seljak, 2004). The II power spectrum is modelled as

$$C^{\rm II}_{\ell,ij} = \int_0^{\chi_{\rm H}} {\rm d}\chi \, \frac{n_i(\chi)\, n_j(\chi)}{\chi^2} \, P_\delta(k;\chi)\, F^2(\chi) \tag{6.12}$$

where $F(\chi) = C_1 \rho_{\rm crit} \Omega_m / D_+(\chi)$. C_1 is a constant given by $5 \times 10^{-14}\, h^{-2} {\rm M}_\odot^{-1} {\rm Mpc}^3$, $D_+(\chi)$ is the linear growth factor normalised to unity today and $\rho_{\rm crit}$ is the critical density of the Universe today. On the other hand, the GI power spectrum is modelled as

$$C^{\mathrm{GI}}_{\ell,ij} = \int_0^{\chi_{\mathrm{H}}} \mathrm{d}\chi \, \frac{w_i(\chi) n_j(\chi) + w_j(\chi) n_i(\chi)}{\chi^2} \, P_\delta(k;\chi) \, F(k;\chi). \tag{6.13}$$

Note that A_{IA} in Equation 6.11 is a free amplitude parameter, which is marginalised over in the sampling procedure. Taking into account the two effects due to intrinsic alignment, the shear correlation function can be written as

$$\xi^{\mathrm{tot}}_{\pm} = \xi^{EE}_{\pm} + A^2_{IA} \xi^{II}_{\pm} - A_{IA} \xi^{GI}_{\pm}. \tag{6.14}$$

Another possible source of theoretical systematics is baryon feedback. The latter causes a reduction in the non-linear matter power spectrum at small scales. In the analysis of KiDS-450 data (Köhlinger et al., 2017), baryon feedback was modelled using fitting formulae as derived by Harnois-Déraps et al. (2015). However, it is worth highlighting that these processes are not very well understood, and hydrodynamical simulations were not strictly carried out to match them to weak lensing observations but rather to understand the processes behind galaxy formation and feedback effects.

In this work, following the same approach used by Hildebrandt et al. (2020) for a comparative study, we use HMCode (Mead et al., 2015) where the halo bloating parameter, $\eta = 0.98 - 0.12 c_{\mathrm{min}}$ changes the halo density profile. $c_{\mathrm{min}} \in [2.0, 3.13]$ is a free parameter marginalised over in the likelihood analysis. A value of $c_{\mathrm{min}} = 2.0$ corresponds to an extreme feedback model (AGN) while $c_{\mathrm{min}} = 3.13$ corresponds to dark matter only.

In addition to the above nuisance parameters, two additional parameters, δc and A_c, were introduced to account for the additive shear bias and position-dependent additive bias pattern, respectively. In short, the set of nuisance parameters, which we denote as Ψ is:

$$\Psi = [A_{IA}, c_{\mathrm{min}}, \delta c, A_c]. \tag{6.15}$$

The new $n(z)$ method developed in this work is coupled with the KV-450 likelihood, and both sets of cosmological, Φ and nuisance, Ψ parameters are inferred.

6.3 Results

6.3.1 Simulation

As the simulated catalogue contains the true redshift distribution, one can compare this directly to the $n(z)$ samples gained from the BHM model. For the full survey distribution, 50000 samples (though a full convergence analysis may show that even less is required) were collected on each source's z,t,m cell allocations. Weights were placed on each source

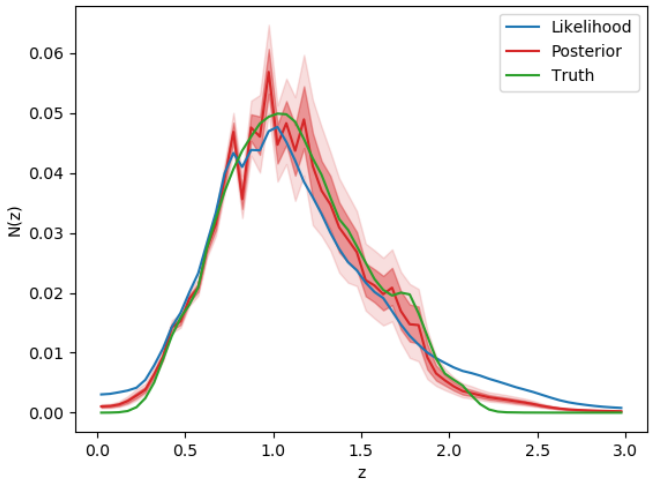


Figure 6.3: The weighted $N(z)$ of the entire survey with the true distribution, posterior distribution and the stacked posterior in green, red and blue, respectively. The posterior is split into 3 tones from deepest to lightest representing the mean, 68% uncertainty and 95% uncertainty distribution.

before totalling the number of sources per z-bin. Figure 6.3 shows the posterior distribution (red) against the true distribution (green). A comparison of mean redshifts shows a much more accurate value than predicted by the stacked posterior and the inclusion of particular uncertainties.

Furthermore, with the inclusion of the BPZ results, one can consider this under the circumstances of tomographic bins with the mock catalogue galaxies separated by peak posterior estimation into the same bins used by KV-450. The BHM model was applied to these smaller samples, as seen in Figure 6.4 and the stacked posterior approach for comparison. Figure 6.4 illustrates the superiority of the BHM model over the obsolete stack likelihood, which suffers from broad distributions and secondary peaks due to degeneracies.

Quantitatively the average χ^2 across the 5 tomographic bins is calculated to be $\bar{\chi}^2 = 102.13$ and $\bar{\chi}^2 = 18.22$ for the stacked posterior method and BHM model respectively. Thus the BHM model shows significant improvement. It should also be noted that both approaches perform worse in the 5^{th} bin as opposed to any other bin (with stacked posterior and BHM having a 5^{th} bin χ^2 of $\chi^2_{5^{\text{th}}} = 307.80$ and $\chi^2_{5^{\text{th}}} = 33.74$ respectively). A possible consequence is that the photometric data is less clear at higher redshifts where more filters would be used. Furthermore, the bias of both approaches (defined as $b(\hat{\theta}) = E_\theta[\hat{\theta} - \theta]$ where $\hat{\theta}$ is the estimated value and θ the true value) is much more comparable but still significantly low being $b = -0.012$ and $b = -0.010$ for stacked posterior and BHM respectively.

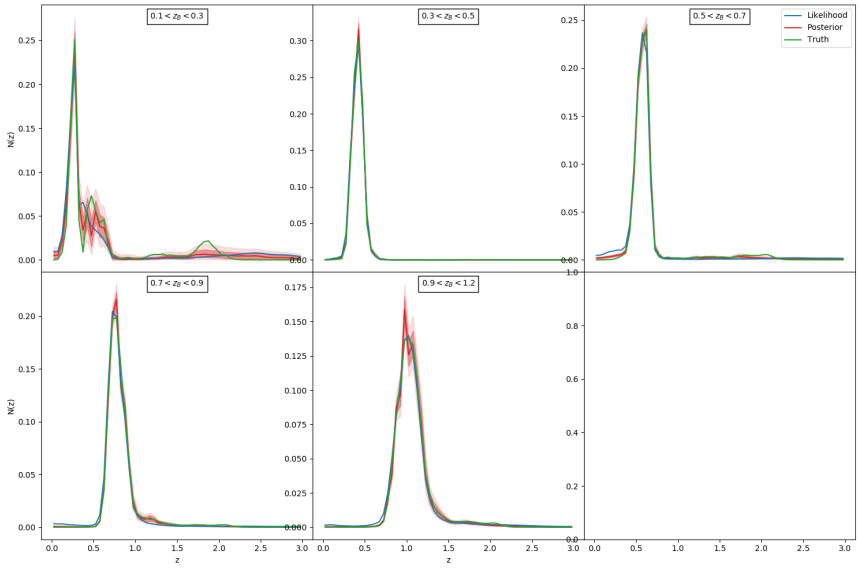


Figure 6.4: Redshift distributions using simulated data with marginalisation over the reference magnitude and type over 5 tomographic bins. The green histogram shows the true input distributions. While the distributions obtained with the inferred parameters f_{ijk} are shown in red. The $n(z)$ gained from the stacked posterior is also included (blue). The posterior is split into 3 tones from deepest to lightest representing the mean, 68% uncertainty and 95% uncertainty distribution.

Table 6.1: The average redshift and its variance for every redshift bin using the 2 models, the stacked posterior (SP) and the BHM model of this paper.

Model	z_B range	$\Delta\mu_z$	$\Delta\sigma_z^2$	ΔS_z
BHM	Full Survey	0.004	-0.003	-0.063
	$0.1 < z_B \leq 0.3$	0.003	0.015	-0.119
	$0.3 < z_B \leq 0.5$	-0.002	-0.0005	-0.601
	$0.5 < z_B \leq 0.7$	0.004	-0.003	0.127
	$0.7 < z_B \leq 0.9$	0.001	-0.002	-0.050
	$0.9 < z_B \leq 1.2$	0.005	-0.003	-0.063
SP	Full Survey	-0.011	0.035	0.509
	$0.1 < z_B \leq 0.3$	0.004	0.213	0.518
	$0.3 < z_B \leq 0.5$	0.004	0.003	3.904
	$0.5 < z_B \leq 0.7$	0.014	0.076	0.544
	$0.7 < z_B \leq 0.9$	0.012	0.035	0.825
	$0.9 < z_B \leq 1.2$	0.004	0.035	0.509

Although it is not visually clear from the tomographic bins where the added accuracy comes from, to truly appreciate the differences between stack likelihood results and the BHM, consider Figure 6.5. In this case, the entire survey is again observed with only the templates t marginalised instead of the templates and magnitudes. Under these conditions, we can isolate some of the data sections of the lowest signal-to-noise. As stated in the 6.1, the simulated data was created to lower signal to noise as the magnitude was increased. This demonstrates that we have less information on the faintest of sources. In the stacked posterior, the lowering of signal to noise is shown through the broadening of the peak likelihoods as we move to fainter magnitude bins. This broadening is an inaccurate survey representation and a complete departure from the actual distribution (green). One may argue that this is what *should* happen; as one has less information, so does there surety that a galaxy is at any one redshift; thus, the uncertainty grows with the broadening of the likelihood. Unfortunately, this is not the case and confusion about the differences between a single galaxy estimation and a survey estimation. If the last magnitude bin represented a single galaxy's likelihood, it would show the proper growth of uncertainties. However, being a survey estimation, the stacked posterior is more akin to a single-point estimation. There is no PDF of the PDF; thus, with a lack of uncertainties, the stack likelihood is to say that the broad distribution *is* the actual distribution. Contrast this with our posterior, which, although it suffers from a small extent of broadening, is mitigated through the correct growing of uncertainty as shown through the 68% and 95% uncertainty bands.

Other advantages can also be visible: stacked posteriors average out and smooth down a distribution acceptable in high signal-to-noise areas. However, as one can see in the final 2 magnitude bins, this has the negative effect of completely losing the secondary peak and showing an inability to move away from the uniform prior, as demonstrated through the probabilities never approaching zero for regions where the actual distribution shows no galaxy is to exist. Whereas with the BHM model, the secondary peaks continue to be inferred even if they are slightly dilute by the growing uncertainty (as shown by the bump as opposed to a peak in the last magnitude bin). It may seem paradoxical how a method that actively suppresses secondary peaks could lead to the survival of one, yet this is once again a confusion of individual galaxies to a complete survey. Individual galaxies have secondary (or even primary) peaks suppressed through the distribution, favouring one over the other, thus breaking the template degeneracy problem. However, on the scale of the entire survey, the Gibbs sampler is actively sampling the distribution for features that can inform the individual galaxies. Such features may include the prominence of galaxies at a second redshift peak. The same can be said for regions of low galaxy count; due to broad individual likelihoods, the raising of the posterior away from zero cannot be avoided entirely, but the damage is

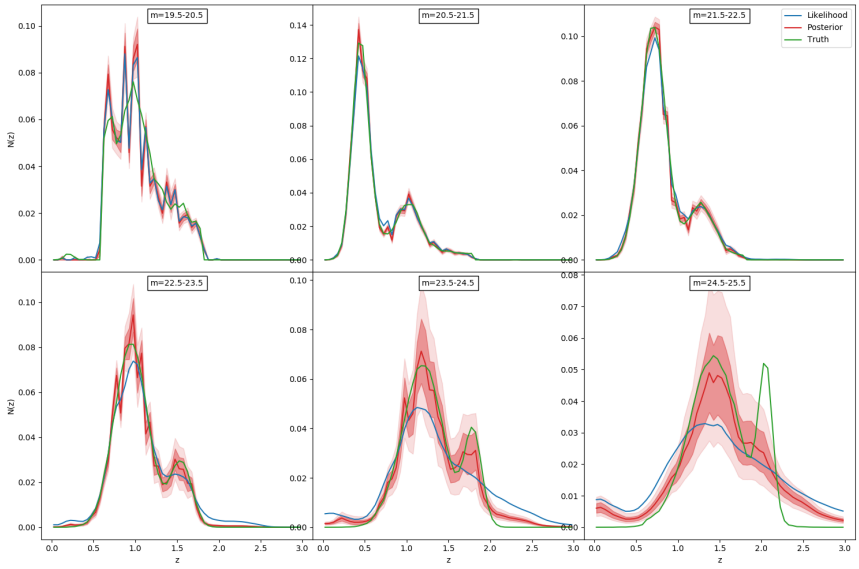


Figure 6.5: The weighted $N(z,m)$ of the entire survey with the true distribution, posterior distribution, and the stacked posterior in green, red and blue, respectively, in this case, the templates t have been marginalised out. The posterior is split into 3 tones from deepest to lightest representing the mean, 68% uncertainty and 95% uncertainty distribution.

suppressed by distribution as a whole, not expecting galaxies to be in this region. For the best example, consider magnitude bins $m = 23.5 - 24.5$ where the stack likelihood has a high probability at high redshift, whereas the posterior lower much quicker in these regions.

6.3.2 KiDS + VIKING 450

A randomly picked set of 100,000 galaxies from the KiDS Survey were analysed using our BHM method to infer $n(z)$ samples. The weighted $n(z)$ for the 5 tomographic bins can be seen in Figure 6.6. Although one cannot make quantitative assessments in the same way seen for the simulated data due to the lack of true distribution, one can still make a qualitative comparison to the results based on the same data for stacked posteriors and DIR approaches. One of the main features is that at low tomographic bins, much like the stacked posterior, the posterior cannot find the tail of high redshift sources appearing in the DIR analysis. This agreement with stacked posterior is expected as one can see from the simulated results that low tomographic bins correspond with low magnitudes and thus high signal to noise which would lead to sharp likelihood distributions as we have small errors on their fluxes. The distribution would not easily sway these sharp likelihoods as a whole being at a different redshift and thus the lack of difference between the 2. The discrepancy from the DIR result is troubling only if we believe DIR to be the truth. If we instead look toward our

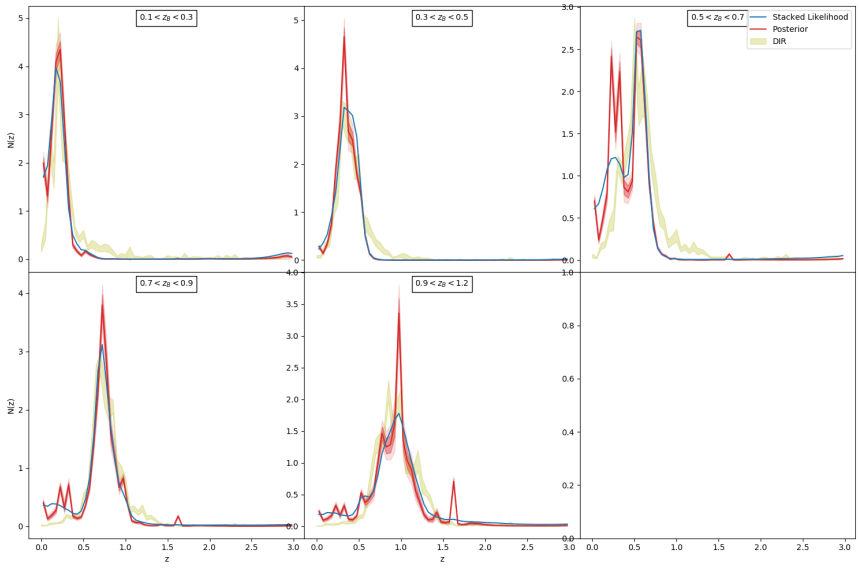


Figure 6.6: Redshift distributions using KV-450 data with marginalisation over the reference magnitude and type over 5 tomographic bins. The cream histogram shows the results from KiDS DIR samples used for their analysis. The stacked posterior is in blue. While the distributions obtained with the inferred parameters f_{ijk} are shown in red. The $n(z)$ gained from the stacked posterior is also included (blue). The posterior is split into 3 tones from deepest to lightest representing the mean, 68% uncertainty and 95% uncertainty distribution.

simulated data, then our low tomographic assessments are correct, and in 4.3.1 we covered the situations in which the stacked posterior can also be reliable, a situation that may cover the low tomographic bins. The other prominent feature is the 'spikiness' of the distribution, which may lead to artificially large peaks, as seen in almost every tomographic bin. Reasons and potential remedies for this spikiness will be addressed below.

To examine the issue of template sensitivity, 3 groups of randomly chosen 50 template sets and 3 groups of randomly chosen 100 template sets were used to produce redshift distributions for the 5 tomographic bins. The median redshift of the distributions was then compared to the PCA result to prove consistency. The median redshift was chosen as the comparison statistic due to its influence over the weak lensing power spectrum; thus, any variability in its value may affect cosmological results. For example, from Figure 6.7 *left*, we can see that when choosing 50 random templates, there is too much variability within the median redshift result. However, the issue is no longer present when one chooses a set of 100. Showing 100 templates to be an adequate number leads to template insensitivity. Furthermore, when compared to the 50 templates present in the PCA analysis, one can see that the PCA result is in complete agreement with the 100 template results showing

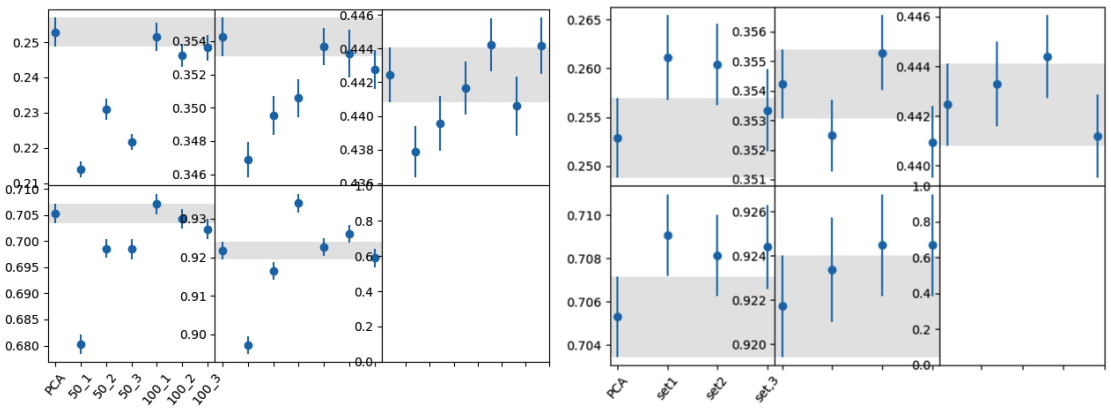


Figure 6.7: *left*: The median redshift based on the redshift distribution of the KiDS450 -Viking data and their variance compared to 6 random template sets (3 containing 50 templates and 3 containing 100 templates), these sets are compared to the same result gathered from PCA analysis (grey). *right*: The median Redshift based on the redshift distribution of the KiDS450 -Viking data and their variance compared to 3 randomly chosen sets of 100,000 galaxies, all of which used the PCA template set. The set used for cosmological analyses is highlighted in grey.

that the reduced number of 50 templates within the PCA set still contains similar levels of completeness when compared to the 100 template counterparts.

A similar analysis was taken in the decision to use only 100000 galaxies for the redshift analysis, as can be seen in Figure 6.7 *right* when we look at the median redshift using the PCA templates set on different randomly chosen sets of galaxies (all though random no set contained the same galaxies as another set). Again, we find complete agreement meaning 100,000 galaxies is an adequate number for the correct analysis while mentioning efficiency.

For the cosmological analysis, we randomly picked sets of 10^5 galaxies and inferred the $n(z)$ tomographic redshift distributions. For consistency check, we repeat the experiments four times and store all the samples of the heights of the redshift distributions. We also generate a fifth set by combining all the samples for each independent set. We will henceforth refer to the fifth set as the “Combined set". We then take the second half of the MCMC chains and use an appropriate thinning factor to generate 5000 and 15,000 $n(z)$ samples. We then sample the posterior using each set, with the 5000 and 15,000 samples separately. We did not find any significant change in the values of the parameters inferred, and throughout this work, we report only the results obtained with the 15,000 $n(z)$ samples. Note that Köhlinger et al. (2017) marginalised over the $n(z)$ uncertainties using 1000 samples of $n(z)$. In some analyses, only the mean of the $n(z)$ is used in the sampling procedure. See, for example, the work by Hildebrandt et al. (2020).

In Figure 6.8, we show the full 1D and 2D marginalised posterior distribution of all the cosmological and nuisance parameters. The inner and outer contours are plotted at 68% and

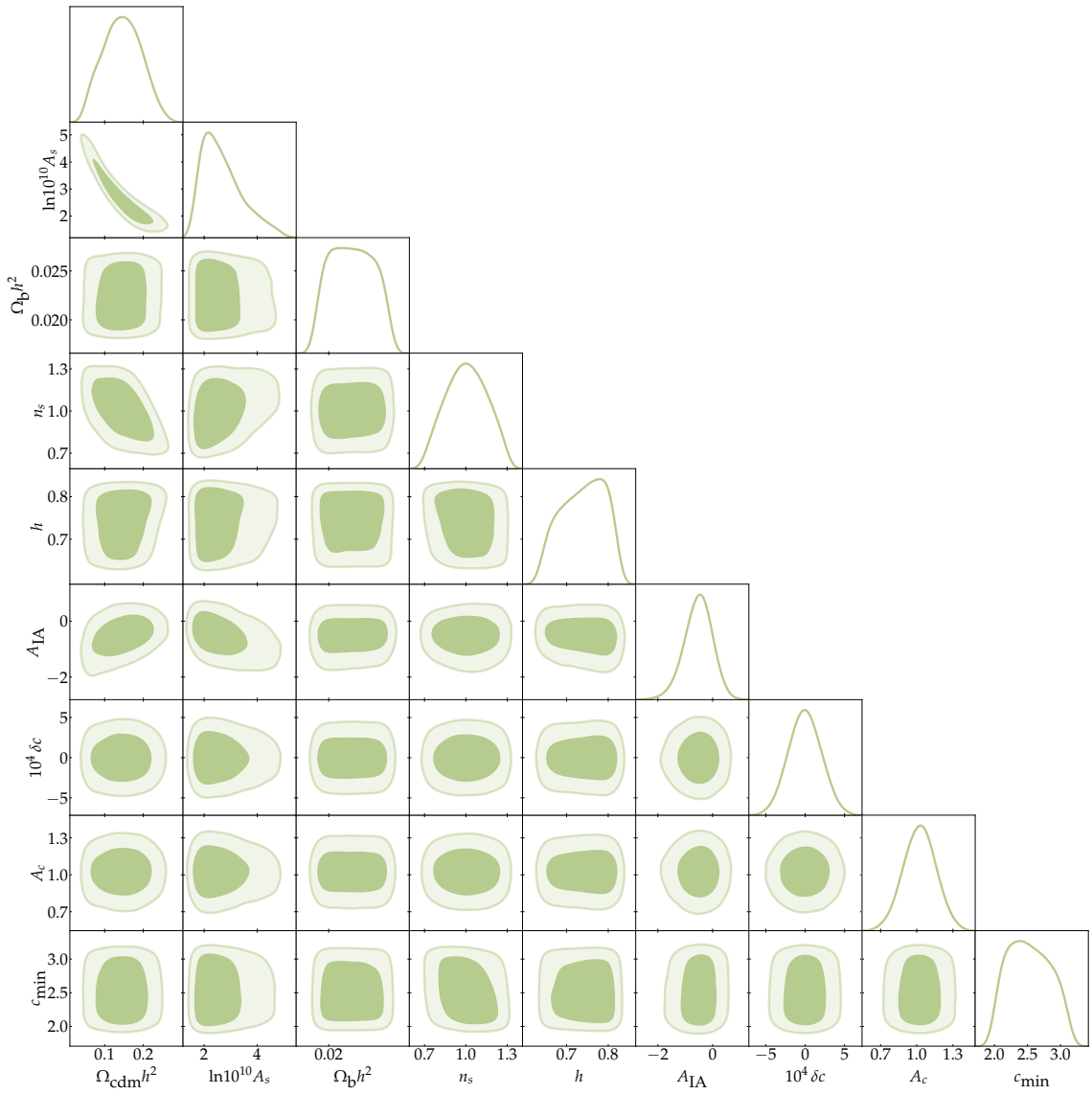


Figure 6.8: The full marginalised 1D and 2D posterior distribution of both sets of cosmological and nuisance parameters. The inner and outer contours correspond to the 68% and 95% credible intervals. A top-hat prior is assumed for all parameters except for δc and A_c, for which Gaussian priors are adopted.

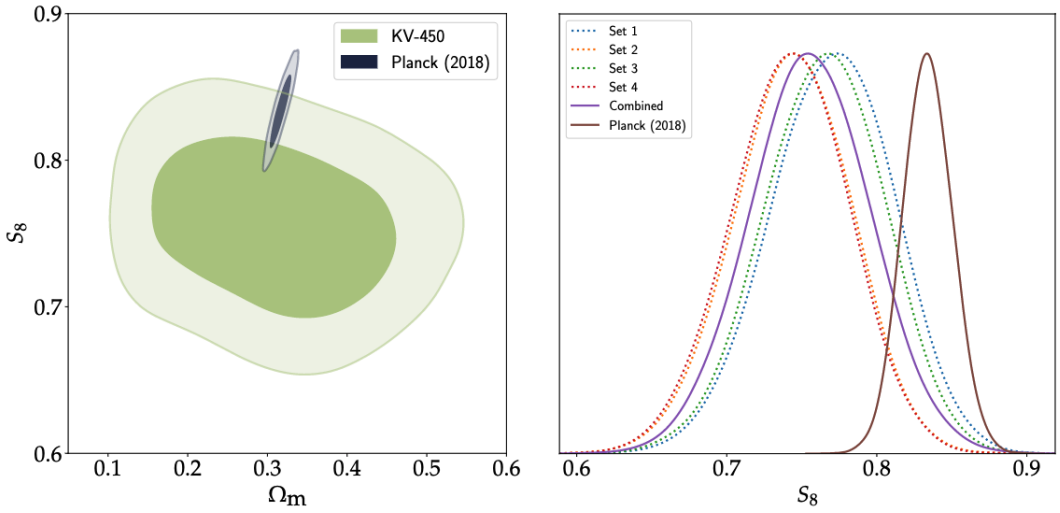


Figure 6.9: *left*: The marginalised posterior distribution in the $\Omega_m - S_8$ plane for the KV-450 analysis (in green) and Planck (in blue). The inner and outer contours correspond to the 68% and 95% credible intervals, respectively.*right*: The marginalised posterior distribution for the S_8 parameter in various experiments. The broken curves correspond to the results obtained when different sets of $n(z)$ are used, and the purple curve shows the results when the sets are combined. The distribution of S_8, as obtained by the Planck 2018 results (Aghanim et al., 2020) is also plotted in brown.

95% credible intervals, respectively. A flat uniform prior is assumed for most parameters except for δc and A_c for which Gaussian priors are assumed. The mean and 1σ credible intervals are also reported in Table 6.2. We find a slightly higher value in the value of $S_8 = 0.756^{+0.039}_{-0.039}$ compared to the original KV-450 analysis.

In Figure 6.9 *left*, we compare the $S_8 - \Omega_m$ plane as obtained in this work using the combined set of $n(z)$ with the one in the latest Planck 2018 results. Similarly, in Figure 6.9*right*, we focus only on the 1D marginalised posterior distribution of S_8 and compare the results obtained with each set (1, 2, 3, 4, "combined") with the Planck 2018 result. As expected, the posterior due to the combined set is roughly centred around the other posterior distributions for KV-450.

To quantify the statistical distance in the value of S_8 among the different experiments performed in this work, we choose to compute the Kullback-Leibler divergence, which is given by

$$D_{\mathrm{KL}}(P||Q) = \int p \log\left(\frac{p}{q}\right) dx, \tag{6.16}$$

where q is the reference distribution. In the first case, we take the distribution due to the combined set as the reference distribution, and the KL-divergence measures are 0.0663, 0.0451, 0.0259 and 0.0578 due to Sets 1, 2, 3 and 4, respectively. On the other hand, we take

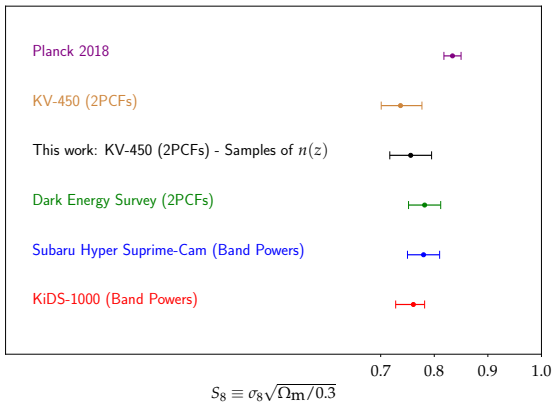


Figure 6.10: The inferred values of S_8 in different surveys and collaborations. The third row corresponds to the value obtained in this work.

the distribution of S_8 inferred by Aghanim et al. (2020) as the reference distribution, and the KL-divergence measures are 10.2, 16.4, 11.4, 16.8 and 13.7 due to Sets 1, 2, 3, 4 and Combined respectively. By visually inspecting Figure 6.9 *right*, these values are aligned with the positions of the different distributions.

Moreover, in Figure 6.10, our result is compared to other surveys and obtained with the DIR method in the original KV-450 analysis. In the latter, the "marginalisation" over the $n(z)$ uncertainty is carried out by including uncorrelated shifts along the redshift axis for each tomographic redshift bin. Hence, the original KV-450 analysis has five additional nuisance parameters. Our method differs because samples of the $n(z)$ are first generated using the Bayesian Hierarchical model. These samples are then used to marginalise the $n(z)$ uncertainty in the likelihood analysis. While most of the results from weak lensing analyses tend to agree with each other, there remains the tension with the value of S_8 from Planck, although, in this work, this tension is further alleviated with our method for determining the $n(z)$ distributions.

We have presented a novel approach for generating redshift distribution samples under a Bayesian Hierarchical framework without needing spectroscopic re-calibrations, as with modern redshift methods. The approach was tested using the simulated data specifically designed to highlight the weaknesses of current photometric techniques and was found to be robust in capturing accurate redshift distributions even in areas of high magnitude and low signal to noise, the area of most concern for weak lensing surveys. Furthermore, by comparing the distribution to that gained from the stacked posterior method, we have shown that the BHM model is more accurate in estimating redshift distribution in photometric data analysis.

Table 6.2: Result for KV-450 analysis with new $n(z)$ redshift distribution.

Parameter	**Symbol**	**Value**
CDM density	$\Omega_{\rm cdm}h^2$	$0.146^{+0.050}_{-0.051}$
Scalar spectrum amplitude	$\ln 10^{10}A_s$	$2.700^{+0.785}_{-0.744}$
Baryon density	$\Omega_b h^2$	$0.022^{+0.002}_{-0.002}$
Scalar spectral index	n_s	$1.005^{+0.155}_{-0.154}$
Hubble parameter	h	$0.739^{+0.056}_{-0.058}$
IA amplitude	$A_{\rm IA}$	$-0.530^{+0.489}_{-0.490}$
Baryon feedback amplitude	$c_{\rm min}$	$2.529^{+0.352}_{-0.342}$
Constant $c-$term offset	$10^4\delta c$	$-0.026^{+2.035}_{-2.031}$
2D $c-$term amplitude	A_c	$1.024^{+0.131}_{-0.133}$
Matter density	Ω_m	$0.311^{+0.095}_{-0.099}$
$\sigma_8\sqrt{\Omega_m/0.3}$	S_8	$0.756^{+0.039}_{-0.039}$

We have also used the samples of the $n(z)$ redshift distribution gathered from using KV-450 data for constraining cosmological parameters with the KV-450 data products. We keep all configurations fixed and replace only the original $n(z)$ redshift distributions, generated using the DIR technique, with the samples obtained in this work. Note that we marginalise over the $n(z)$ by drawing random samples of $n(z)$ at each likelihood evaluation, whilst the original KV-450 analysis employs offsets in the $n(z)$ distribution along the redshift axis. As a result, our analysis yields a slightly higher value of $S_8 = 0.756^{+0.039}_{-0.039}$ compared to the original KV-450 analysis, which reports a value of $S_8 = 0.737^{+0.040}_{-0.036}$.

The tension in S_8 between weak lensing analyses and the Planck analyses has been a topic of heated debate, and in this work, we show that using Bayesian Hierarchical modelling and effectively marginalising over the $n(z)$ uncertainty leads to a reduction in the tension between the two probes. Identifying the exact source of the discrepancy in S_8 remains an open research question. One school of thought might argue that modelling the systematic in either analysis is a complex problem, while another posits that there can also be other Physics models to explain the underlying data. Much is left open for improvement, especially on our code's computational aspect, which limited us to a small population and template size. Furthermore, the spikiness of the BHM distribution could be investigated further to clarify their existence. However, further refinement of personalising inference for each survey through template choice (or an external distribution error to approximate template sensitivity) and relevant priors to the distribution could be included as all priors of this work were considered uniform.

As well as the implementation and relevance of selection effects. When derived using this method, the $n(z)$ distributions should be useful for parameter estimation for future surveys such as DES, LSST and Euclid, of which deep and numeral spectroscopic data will not be available.

Chapter 7

KiDS-1000

Through the previous chapter, we were able to show the benefits of using a Bayesian hierarchical approach to redshift distribution inference and how samples from said distribution can be used for cosmological analysis. However, improvements could be made and will be required to reach the precision and accuracy needed for the next generation of weak lensing surveys. With that in mind, this chapter covers our unfinished work into possible improvements one could make and apply our method to one of the most recent weak lensing surveys, KiDS-1000.

7.1 Regularisation

As stated in 6.3, there is sometimes a 'spiky' quality to the redshift distributions inferred from the BHM, a quality one presumes is a fault of the method and not necessarily a true reflection of the redshift distribution. Below are our first attempts to understand and potentially remedy this spikiness.

One of the main advantages of the original theory was the suppression of secondary peaks through a process already mentioned as 'Bayesian Shrinkage'. A source of quite ambiguous z, m, t can be allocated to a cell with very little flexibility, especially if other members of its population highly favour that cell. This helps develop the 2 original effects seen. Firstly the model can infer set (m,t) or (m,z) relations without the need for specific non-uniform priors, as with BPZ. Secondly, it helps remove the tails one usually associates with stacked likelihoods as the ambiguous source can still have highly peaked posteriors even with broad likelihoods. Both effects and their advantages can be seen at work in Figure 6.5.

However, as valuable as the Bayesian Shrinkage is, some issues develop under modern weak lensing surveys, specifically in the case of high populations with low S/N. Under these conditions, the uniform prior that suggests an equal amount of sources per cell has minimal

impact on the possible Dirichlet distributions. The variance in possible distributions from equation 5.5 is minimised. Furthermore, a large population of low S/N sources will trend towards a small set of z,t,m cells as long as a tiny amount of higher S/N sources of the same population have less overall likelihood in this region regardless of if that region represents the truth.

Thus, a large population of low S/N sources can have a very peaked redshift distribution with minimal variance. Lastly, due to the Dirichlet distributions having every cell anti-correlated with every other cell (a direct consequence of $\sum_{ijk} f_{ijk} = 1$ so if one cell rises, the rest must fall) and the likelihood functions having a positive correlation with nearest neighbours in redshift (as the likelihood is Gaussian) the combination of the 2 means nearest neighbours in z-cells anti-correlate leading to a lack of smoothness one expects to find in redshift distributions.

In this sense, the model has proscribed a distribution quality that may bias results and something that will only grow worse with a higher population and lower S/N data. To remedy this issue, regularisation may be used to ensure smoothness in the redshift distribution. This same regularisation was not required for the m,t bins for different reasons. Although neighbouring m-bins would have the same correlation issues mentioned above, the likelihood distributions in m are exceptionally constrained due to the wealth of information from the flux data. This favours the likelihood peaks over any peaks suggested by the population as a whole, negating the issue. While t-bins are allocated randomly, so have no association with their t-neighbours.

The simplest method would be to smooth the distribution post-production through a convolution operator applied to the distribution in the z-axis after all samples have been acquired. This method was swiftly rejected due to not being a true reflection of the inferred samples and also that the damage has already been done as a source have already located in areas they may not have been in at all thus smoothing them to nearest neighbours is just spreading the problem not solving it. An improvement on this regularisation attempt was a 3-point moving average applied to f_{ijk} in the z-direction after every Gibbs sample. This is to say that the distribution is smoothed between equations 5.5 and 5.4. This means that the following collection of n_{ijk} samples is informed by the fact that the underlying distribution f_{ijk} is smooth, which makes sense. We expect the underlying distribution to be smooth, but not necessarily the number counts of the sources themselves. Aware that the use of a 3-point moving average is a departure from the Gibbs-sampling method and courses of the chance of over-smoothing (as it takes place in every sample) in a potentially non-smooth redshift population, we chose to weight our 3-point moving average to the ratio of $[1,20,1]$ this was the ratio that had a minor impact on the distribution wiliest still having the desired effect.

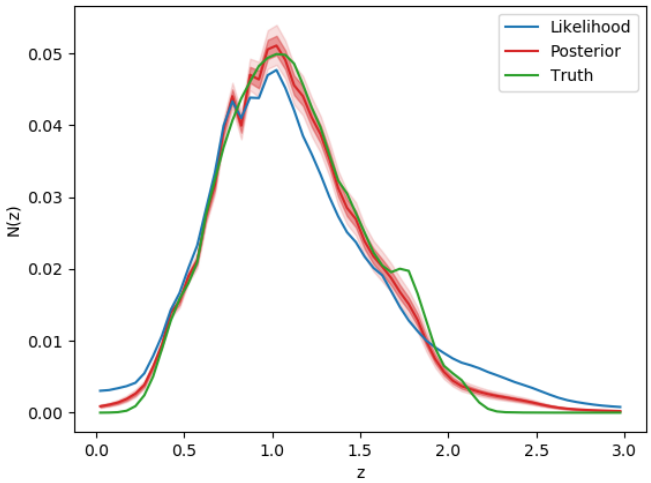


Figure 7.1: The weighted $N(z)$ of the entire survey with the true distribution, posterior distribution and the stacked likelihood in green, red and blue, respectively. The posterior is split into 3 tones from deepest to lightest representing the mean, 68% uncertainty and 95% uncertainty distribution. The version has samples based on the smoothed variant.

In Figure 7.1, we can see the full distribution results, which can be directly compared to Figure 6.3 as there based on the same simulated data. Again, one can see a reduction in spikiness, and at redshifts of < 1.7, the inferred distribution is both more accurate (as the mean is closer to the actual distribution) and more precise (the 68% and 95% error bars are smaller) but the smoothed samples do a worse job of identifying the secondary peak.

Even with the successes, we ultimately believe this idea requires more research the sheer ad hoc nature to which we apply a smoothing function of chosen weight removes the principled nature of the BHM approach, i.e. the smoothness of the function needs to be inherently baked into the model and be physically motivated. Also, the success may result from the smooth simulated model instead of the model inferring this distribution any better. For baking the smoothness, one can use additional priors. For example, a higher number of uniform α values in prior would imply a more substantial uniformity to the distribution preventing peaks from becoming too high. With that said, a consequence would be the broadening of distributions as seen with the stacked likelihood. Alternatively, one may want to consider using a different kind of Dirichlet function, such as a generalised Dirichlet function (or 'hyperdirichlet') (Hankin, 2010) to which control over correlation and meta-structure may be possible. With that said, this would introduce multiple new parameters and consider the computational difficulties one has already had. This may not be practically possible.

7.2 Gaussian Processes

Regarding computational difficulties, cosmology was one area of an extreme bottleneck during the analysis of Chapter 6. To compute cosmological parameter inference, we require a forward model in the form of CLASS. This is already very time-consuming due to the $n(n+1)/2$ auto- and cross-correlations needed at every probed angular scale where n is the number of tomographic bins. An issue is exacerbated by adding $n(n+1)$ power spectra if intrinsic alignments are included. The issue is just manageable in its current state, as demonstrated through the ability to gather cosmological parameters as seen above; however, this will not be the case going forward in new weak lensing surveys with multi-year estimates being made for the time scale of weak lensing analysis for future surveys based on current methods (Heitmann et al., 2009).

With this in mind, we plan to future-proof the practical use of our $n(z)$ samples by showing their ability to be used with an emulator, a device that replaces the full forward model simulation with an approximate mathematical function. Thus future studies of our BHM model will make its cosmological inferences using the Gaussian processes emulator as created by Mootoovaloo et al. (2020). In this work, 25 Gaussian processors (one for each band power) are formed by training them on the band powers created on a set of parameters (training points) for CLASS. The advantage of Gaussian processes is that it produces a mean value and a variance, allowing for total uncertainties to be accounted for, just as we saw in the previous BHM. As one can see from Figure 7.2 the processor does an outstanding job at replicating the parameter inference one would have got would a CLASS-based forward model, both through using just the mean or including the variance on the emulations. This comes with the added benefit of being at least 20 times faster than the current methods. Work has begun on incorporating the samples; currently, they work using the mean distribution (as in the mean probability in every z-bin within a tomographic bin, not the mean value of the distribution as seen in Figure 6.7) but further work needs to take place for all $n(z)$ samples to be incorporated.

7.3 KiDS-1000

The next set of data we wish to use our current BHM $n(z)$ inferred as well as the Gaussian processes emulator for the cosmology is with the Kilo-Degree surveys 4^{th} data release (Kuijken et al., 2019). Through the use of the square 268-million pixel, CCD mosaic camera OmegaCAM (Kuijken, 2011), and the VIKING survey (Edge et al., 2013) the amount of sky coverage available in all 9 bands *ugriZYJHK* has more than doubled from 450 square degrees

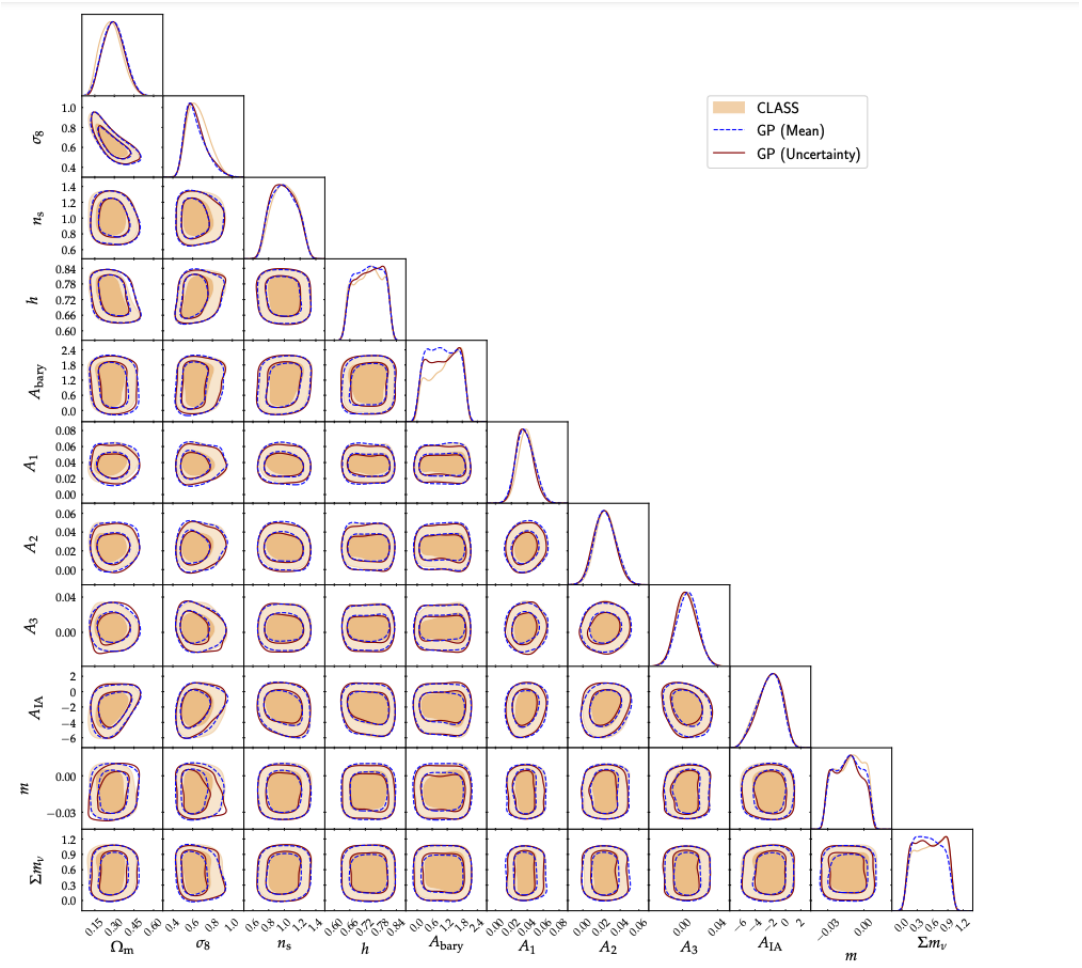


Figure 7.2: The full 1D and 2D marginalised posterior distributions were obtained using three different methods - The one in tan colour corresponds to posterior distributions with the full simulator (CLASS), while the solid brown one corresponds to the Gaussian Process emulator when random functions of the band powers are drawn, marginalising the Gaussian Process uncertainty. The posterior in blue shows, the distributions obtained when only the mean of the Gaussian Process was used in the inference routine. The contours denote the 68% and 95% credible interval, respectively. Note that some parameters are dominated by their respective priors and are not constrained at all. A similar conclusion was drawn by Köhlinger et al. (2017). However, the important point here is that the posterior from the GP is close to that obtained with CLASS Mootoovaloo et al. (2020).

to 1006. The added coverage can be seen visually in Figure 7.3 which all the releases are shown in green, whereas blue represents the new data in the most recent data release and grey the goal of 1350 square degrees in the future. As well as a considerable increase in area, this latest release also represents some technological improvements. One of the main changes is that KiDS are now using the deep, optical, all-sky catalogue of Gaia data release 2 (Brown et al., 2018) to calibrate the observed photometry. This is because each KiDS square degree tile contains 1000s of stars that Gaia observed in broadband filters with accuracy better than 0.01 magnitudes, far beyond the accuracy of KiDS. The colour-colour diagrams of similar regions are compared, and offsets are calculated for KiDS to be in line with Gaia allowing more accuracy in the observed galaxies. The data by which our studies are preoccupied is the GAAP-b, where b are the 9 filter bands. This data column (unlike KiDS-450) is itself not a single observation but an optimum choice between 2 aperture images (1.0" or 0.7"). The following conditions decide the optimum choice:

1. For all bands b calculate the flux error ratios:

$$R_b = \frac{\text{FLUXERR-GAAP-1p0-x}}{\text{FLUXERR-GAAP-0p7-x}}. \tag{7.1}$$

2. If $\max(R_b) \times \min(R_b) < 1$ use the 1.0" image otherwise use the 0.7" image

This ensures that the smaller aperture is used unless the rare circumstance applies (in about 4% of all cases) where the larger aperture gives more minor flux errors. This allows for a second chance for better image quality on the rare occasion when the smaller aperture fails. To use these fluxes, the extinction needs to be subtracted just as we did for KiDS-450. The only difference is that the extinction values are not just interpolated from the extinction maps by Schlegel et al. (1997) but also combined with the conversion coefficients of (Schlafly and Finkbeiner, 2010) for added accuracy. To allocate sources to different tomographic bins, the same BPZ photometric code is used as before, but this time with prior from KV450 as opposed to K450 (Raichoor et al., 2014). The choice was made to change from the default prior as this new one reduces uncertainties and catastrophic errors on faint galaxies; however, it comes with a biasing of bright low-redshift galaxies to slightly higher redshifts, so the results should only be seen as a categorising tool and not an accurate measurement on the redshifts of the sources.

With this in mind, we applied the same BHM code used for the KiDS-450 analysis to our new flux data. Once again, we used 100,000 randomly selected sources and the 50 PCA templates for the best balance in completeness and efficiency. 50,000 samples were selected to ensure convergence, and sources were separated into their tomographic bins based on the BPZ data given by the DR4 release. The results can be seen in Figure 7.4.

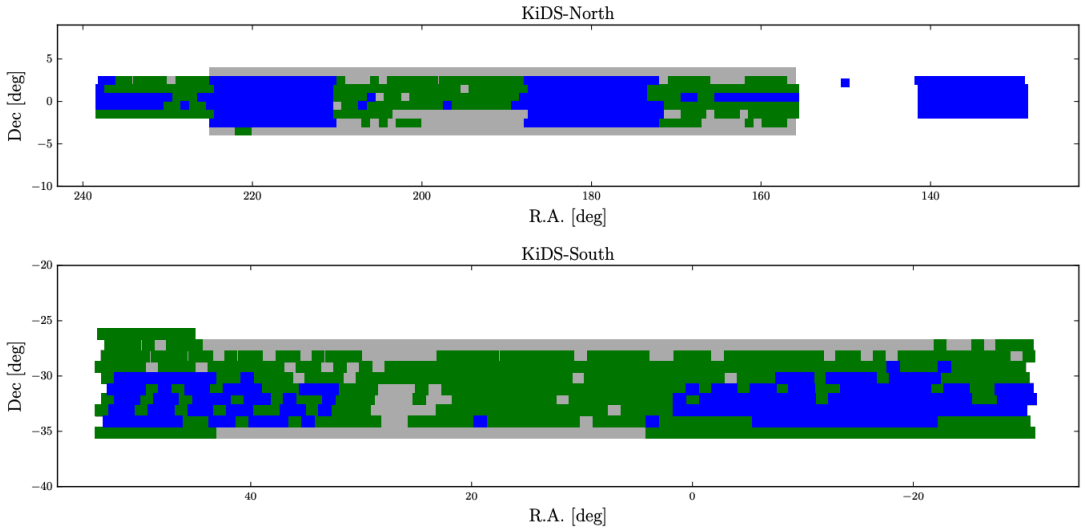


Figure 7.3: The sky coverage of the KiDS survey 4^{th} data release DR4, the blue represents all the accumulated data from the first 3 releases culminating in KiDS-450, whereas the blue represents all the new data. The grey represents the end goal coverage of 1350 square degrees (Kuijken et al., 2019) .

Note that these results are highly preliminary, given that no other improvements to the code were used. However, on first inspection, the results seem promising, with the peak redshifts being more in line with DIR and SOMz than we saw previously. However, there is a glaring issue of very high redshift sources, especially in the first 2 tomographic bins. Simply put, there are sources that KiDS have allocated to low redshift that we have allocated to high. The high redshift sources also appear in the stacked likelihood suggesting this is not an issue with the Gibbs sampling side of the code directly. Instead, this looks like a common degeneracy in which low redshift sources can appear high, a degeneracy not suppressed by our code but seems to be identified and suppressed by KiDS, potentially in their prior change. To investigate the issue further, we repeated the experiment of completing the analysis again for 5 sets of randomly chosen 100,000 sources to see if the results were consistent and not just a consequence of shot noise.

Figure 7.5 suggests that even though the survey size has gone from 13,241,807 sources in KiDS-450 to 21,262,011 sources in KiDS-1000, the median redshifts based on a randomly selected 100,000 is even more consistent than their KiDS-450 counterparts (you may have noticed that the number of sources has not doubled even though the coverage has, this is due to the reduction in sources being used as the culling has taken place that was mentioned in the SOMz analysis 4.3.2). Also, in the first tomographic bin, the median redshift is consistent at $z > 1.2$, much larger than the $0.1 < z < 0.3$ region to categorise.

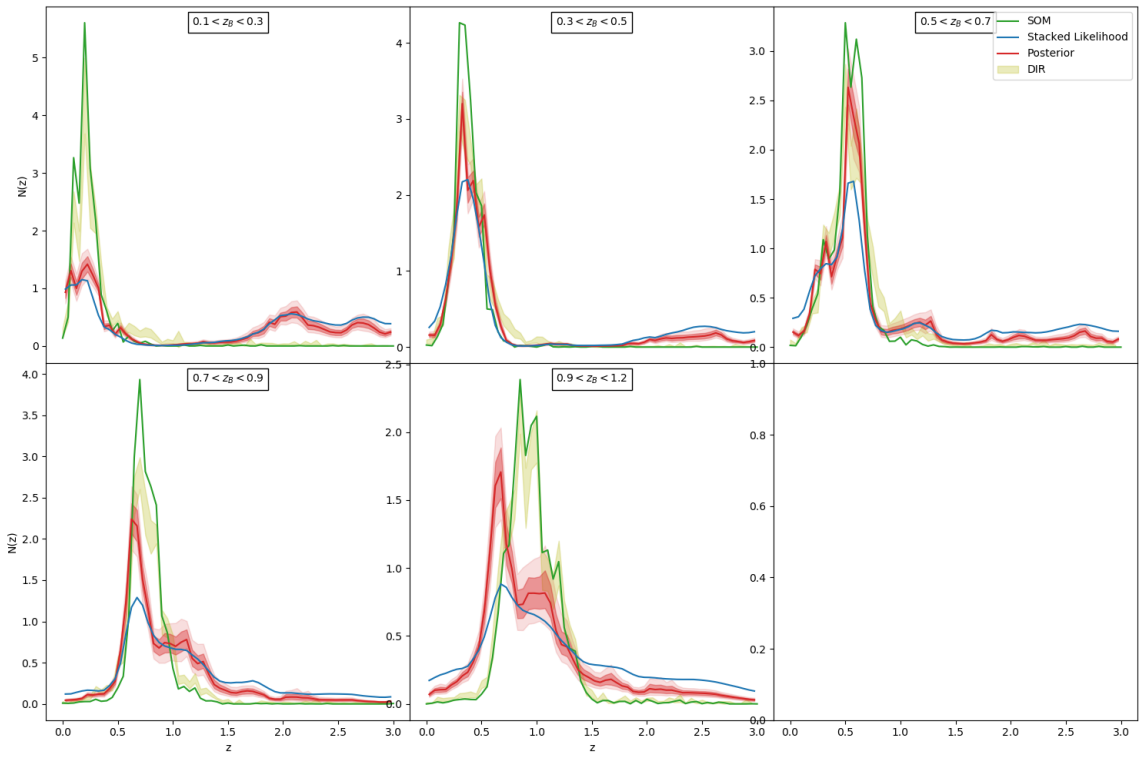


Figure 7.4: Redshift distributions using KiDS-1000 DR4 data release with marginalisation over the reference magnitude and type over 5 tomographic bins. The cream histogram shows the results from KiDS-450 DIR samples, while the green shows the mean redshift distribution of the SOMz work used for their analysis. While the distributions obtained with the inferred parameters f_{ijk} are shown in red. The $n(z)$ gained from the stacked likelihood is also included (blue). The posterior is split into 3 tones from deepest to lightest representing the mean, 68% uncertainty and 95% uncertainty distribution.

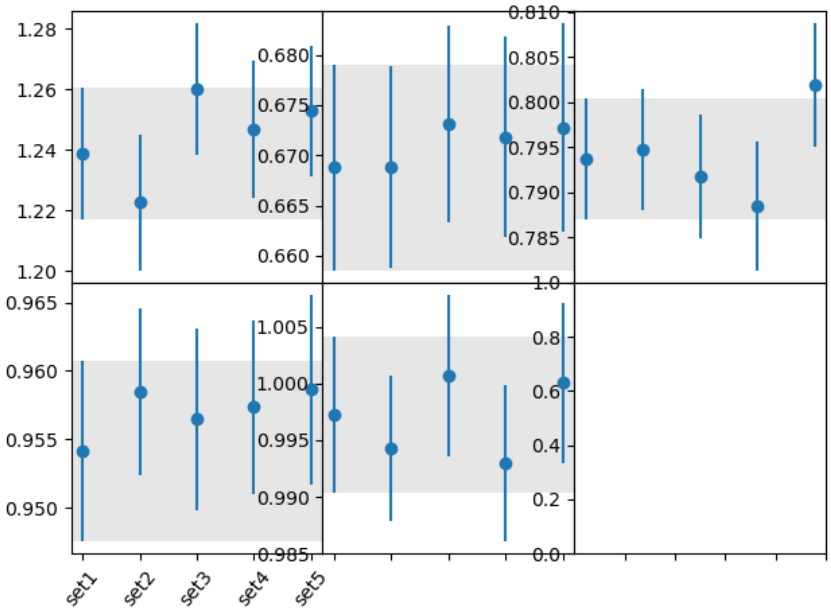


Figure 7.5: The median redshift based on the redshift distribution of the KiDS-1000 DR4 data release and their variance compared to 5 randomly chosen sets of 100,000 galaxies, all of which used the PCA template set. The first set used above is highlighted in grey.

To continue investigating this issue may give us a chance to focus more, such as by categorising our templates to see if certain types of galaxies are appearing at particular redshifts. We may also look into incorporating priors directly so that our analysis does contain physical knowledge of the relationships between magnitude and redshift and does not wholly rely on the method itself to discover said relationships.

Chapter 8

Conclusion

In this thesis, we have demonstrated how Bayesian Hierarchical Models can be applied to new problems in order to gauge accurately and precisely the values and uncertainties of redshift distributions through the use of photometric data alone and thus, without the need for overlapping spectroscopic surveys, these same redshift distributions can be sampled and applied to current cosmological pipelines in order to gather constraints on cosmological parameters.

Chapter 1 delved into how we have those cosmological parameters in the first place. Specifically, we considered the different assumptions and models our parameters rely on, such as General Relativity and our assumption of the cosmological principle. We showed that through these theories and principles, one could use 2-point correlation functions as a primary statistical tool for gathering insight into our known universe. We then consider some of the probes we have for estimating the power spectra that describe these 2-point correlation functions, such as the CMB and Lyα forest. Finally, we also looked into the areas in which our current understanding does not fully describe our universe externally to the standard model in the form of dark energy and dark matter and internally in the form of parameter tensions between different observables.

Chapter 2 looked into one of the probes that currently courses that tension; weak lensing (which currently has a S_8 tension with CMB studies). We saw how through Einstein's equations, Fermat's principle and the Born approximation, one can correctly describe the relationship between an image's distortion and the masses between the emitter and observer that coursed said distortion in the trajectory of a photon. Finally, we tackled the question of how one can determine the level of distortion without access to the actual image through the use of mean convergence and shear across multiple galaxies and how an understanding of that statistical distortion can help infer the matter power spectra and thus constrain values of Ω_m, σ_8 and S_8 as a direct consequence of the first 2 parameters.

Chapter 3 looked into one of the areas of weakness of current weak lensing surveys, which is how one can accurately yet efficiently gauge the redshift of observed galaxies. Through an understanding of a galaxy's origin and morphology, one can see how particular spectral features are expected from the light observed from far away galaxies and how these features are moved and distorted but, importantly, not removed by redshift and thus the difference between the observed and expected spectra can determine a galaxies redshift. We considered the most accurate way to use spectral information for redshift estimation in the form of spectroscopy but rejected it due to the sheer amount of data and lack of quality in that data collected from weak lensing survey, which makes spectroscopy impractical. Instead, we focused on photometric redshifts, a similar method that uses more broad light filters. We look into the different methods used to estimate photometric redshifts from empirical-based methods, which use machine learning techniques, to template-based methods, which use physical modelling and consider the advantages and disadvantages.

Chapter 4 considers how we can alleviate some of the disadvantages of template-based methods by considering a different statistical framework. Up to this point, all template-based methods have used a frequentist approach to gathering statistical insight from data. This approach is described and explained in more detail so that we can contrast with the focus of this chapter which is the Bayesian approach to statistics. First, we derive and explain the benefits of Bayes Theorem and why this requires a different way of thinking and apply the new approach to rudimentary questions. Next, we consider some of the choices one has to make when using Bayesian methods, such as which prior to use and how to best collect samples from the posterior. With this information, we show how Bayesian statistics has been used in the past for photometric redshifts in the form of BPZ. Lastly, we consider the main problem of weak lensing photometric redshifts, which is the redshift distribution of the entire survey, how the method of the past (stacked likelihood) has worked, why in future they will fail, how current methods work but rely on spectroscopic data and how at least in theory a fully Bayesian model can be formed to answer said question.

Chapter 5 takes the issue of inferring the redshift distribution issue through a Bayesian approach from the theoretical to the practical through the work of Leistedt et al. (2016), we demonstrate that through changes to the parameter of interest, one can not only form a Bayesian Hierarchical Model for the relationship between individual galaxies, their photometry and the full redshift distribution but also deal with the requirements of weak lensing surveys such as fainter sources, lower signal to noise and most importantly the weighting of galaxies and splitting of said galaxies into tomographic bins. We show in Chapter 6 through simulation that we can accurately infer the redshift distribution (and its uncertainties) of 5 desperate tomographic bins and demonstrate its superiority over current photometric-based

methods. We then applied the method to the current KiDS-450 weak lensing survey and showed how through sampling the redshift distribution, we could use said samples as inputs for an entire cosmological pipeline leading to a slightly higher value of $S_8 = 0.756^{+0.039}_{-0.039}$ compared to the original KV450 analysis, which reports a value of $S_8 = 0.737^{+0.040}_{-0.036}$. This value reduces the tension between weak lensing and CMB studies showing that the tensions may be due to systematic errors in our redshift estimation as opposed to new physics. Lastly, through Chapter 7 we showed how our method might be improved for more modern surveys such as KiDS-1000 through Gaussian Processes and changes to the prior and displayed preliminary results to which further work needs to occur.

Bibliography

Aalseth, C.E. et al. CoGeNT: A search for low-mass dark matter using p-type point contact germanium detectors. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 88(1):012002, 7 2013. ISSN 15507998. doi: 10.1103/PhysRevD.88.012002. URL https://journals.aps.org/prd/abstract/10.1103/PhysRevD.88.012002.

Aartsen, M.G. et al. IceCube Search for Dark Matter Annihilation in nearby Galaxies and Galaxy Clusters. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 88(12): 41, 7 2013. doi: 10.1103/PhysRevD.88.122001. URL http://arxiv.org/abs/1307.3473http://dx.doi.org/10.1103/PhysRevD.88.122001.

Aartsen, M.G. et al. Search for annihilating dark matter in the Sun with 3 years of IceCube data IceCube Collaboration. *Eur. Phys. J. C*, 77:146–163, 2017. doi: 10.1140/epjc/s10052-017-4689-9.

Abbott, B.P. et al. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116(6):061102, 2 2016. ISSN 10797114. doi: 10.1103/PhysRevLett.116.061102. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.116.061102.

Abdalla, F.B. et al. Photometric redshifts for weak lensing tomography from space: the role of optical and near infrared photometry. *Monthly Notices of the Royal Astronomical Society*, 387(3):969–986, 7 2008. ISSN 00358711. doi: 10.1111/j.1365-2966.2008.13151.x. URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2008.13151.x.

Abdalla, F.B. et al. A comparison of six photometric redshift methods applied to 1.5 million luminous red galaxies. *Monthly Notices of the Royal Astronomical Society*, 417 (3):1891–1903, 11 2011. ISSN 00358711. doi: 10.1111/J.1365-2966.2011.19375.X/2/M{_}MNRAS0417-1891-MU2.GIF. URL https://academic.oup.com/mnras/article/417/3/1891/1089504.

Abolfathi, B. et al. The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. *The Astrophysical Journal Supplement Series*, 235(2):42, 4 2018. ISSN 1538-4365. doi: 10.3847/1538-4365/aa9e8a. URL https://doi.org/10.3847/1538-4365/aa9e8a.

Abruzzo, M.W. and Haiman, Z. The impact of photometric redshift errors on lensing statistics in ray-tracing simulations. *Monthly Notices of the Royal Astronomical Society*, 486(2):2730–2753, 6 2019. ISSN 0035-8711. doi: 10.1093/MNRAS/STZ1016. URL https://academic.oup.com/mnras/article/486/2/2730/5475123.

Adams, W.S. The Relativity Displacement of the Spectral Lines in the Companion of Sirius. *Proceedings of the National Academy of Sciences*, 11(7):382–387, 7 1925. ISSN 0027-8424. doi: 10.1073/pnas.11.7.382. URL https://www.pnas.org/content/11/7/382https://www.pnas.org/content/11/7/382.abstract.

Aghanim, N. et al. Planck 2013 results. XXVII. Doppler boosting of the CMB: Eppur si muove. *Astronomy and Astrophysics*, 571:30, 11 2014. ISSN 14320746. doi: 10.1051/0004-6361/201321556. URL http://www.esa.int/Planck.

Aghanim, N. et al. Planck 2018 results: VI. Cosmological parameters. *Astronomy and Astrophysics*, 641:A6, 9 2020. ISSN 14320746. doi: 10.1051/0004-6361/201833910. URL https://ui.adsabs.harvard.edu/abs/2020A&A...641A...6P/abstract.

Agnese, R. et al. Silicon detector dark matter results from the final exposure of CDMS II. *Physical Review Letters*, 111(25), 12 2013. ISSN 00319007. doi: 10.1103/PhysRevLett.111.251301. URL https://arxiv.org/abs/1304.4279v3.

Ahn, C.P. et al. The ninth data release of the sloan digital sky survey: First spectroscopic data from the sdss-iii baryon oscillation spectroscopic survey. *Astrophysical Journal, Supplement Series*, 203(2):2012, 12 2012. ISSN 00670049. doi: 10.1088/0067-0049/203/2/21. URL https://ui.adsabs.harvard.edu/abs/2012ApJS..203...21A/abstract.

Alam, S.M.K., Bullock, J.S. and Weinberg, D.H. Dark Matter Properties and Halo Central Densities. *The Astrophysical Journal*, 572(1):34–40, 9 2001. URL http://arxiv.org/abs/astro-ph/0109392.

Alsing, J. *Bayesian analysis of weak gravitational lensing*. PhD thesis, Imperial College London, London, 2017. URL https://spiral.imperial.ac.uk/handle/10044/1/44571.

Angloher, G. et al. Results from 730 kg days of the CRESST-II Dark Matter search. *European Physical Journal C*, 72(4):1–22, 4 2012. ISSN 14346052. doi: 10.1140/epjc/s10052-012-1971-8. URL https://link.springer.com/article/10.1140/epjc/s10052-012-1971-8.

ANTARES collaboration et al. Limits on Dark Matter Annihilation in the Sun using the ANTARES Neutrino Telescope. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 759:69–74, 3 2016. doi: 10.1016/j.physletb.2016.05.019. URL http://arxiv.org/abs/1603.02228http://dx.doi.org/10.1016/j.physletb.2016.05.019.

Antilogus, P. et al. The brighter-fatter effect and pixel correlations in CCD sensors. *JInst*, 9(3):C03048, 2014. doi: 10.1088/1748-0221/9/03/C03048. URL https://ui.adsabs.harvard.edu/abs/2014JInst...9C3048A/abstract.

Appelquist, T., Cheng, H.C. and Dobrescu, B.A. Bounds on Universal Extra Dimensions. *Physical Review D*, 64(3), 12 2000. doi: 10.1103/PhysRevD.64.035002. URL http://arxiv.org/abs/hep-ph/0012100http://dx.doi.org/10.1103/PhysRevD.64.035002.

Aramaki, T. et al. Antideuteron Sensitivity for the GAPS Experiment. *Astroparticle Physics*, 74:6–13, 6 2015. doi: 10.1016/j.astropartphys.2015.09.001. URL http://arxiv.org/abs/1506.02513http://dx.doi.org/10.1016/j.astropartphys.2015.09.001.

Arina, C., Del Nobile, E. and Panci, P. Dark matter with pseudoscalar-mediated interactions explains the DAMA signal and the galactic center excess. *Physical Review Letters*, 114(1), 6 2015. ISSN 10797114. doi: 10.1103/PhysRevLett.114.011301.

Arnouts, S. et al. Measuring and modelling the redshift evolution of clustering: The Hubble Deep Field North. *Monthly Notices of the Royal Astronomical Society*, 310(2):540–556, 12 1999. ISSN 00358711. doi: 10.1046/J.1365-8711.1999.02978.X/2/310-2-540-TBL004.JPEG. URL https://academic.oup.com/mnras/article/310/2/540/1048431.

Asencio, E., Banik, I. and Kroupa, P. A massive blow for ΛCDM $-$ the high redshift, mass, and collision velocity of the interacting galaxy cluster El Gordo contradicts concordance cosmology. *Monthly Notices of the Royal Astronomical Society*, 500(4):5249–5267, 12 2020. doi: 10.1093/mnras/staa3441. URL https://arxiv.org/abs/2012.03950v1.

Asgari, M. et al. KiDS-1000 Cosmology: Cosmic shear constraints and comparison between two point statistics. *Astronomy and Astrophysics*, 645, 7 2020. doi: 10.1051/0004-6361/202039070. URL http://arxiv.org/abs/2007.15633http://dx.doi.org/10.1051/0004-6361/202039070.

Asztalos, S.J. et al. Searches for astrophysical and cosmological axions, 10 2006. ISSN 01638998. URL www.annualreviews.org.

Aubourg, E. et al. Cosmological implications of baryon acoustic oscillation (BAO) measurements. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 92(12):9, 11 2014. doi: 10.1103/physrevd.92.123516. URL https://arxiv.org/abs/1411.1074v3.

Audit, E. and Simmons, J.F.L. The use of light polarization for weak-lensing inversions. *Monthly Notices of the Royal Astronomical Society*, 303(1):87–95, 9 1998. doi: 10.1046/j.1365-8711.1999.02226.x. URL http://arxiv.org/abs/astro-ph/9809355http://dx.doi.org/10.1046/j.1365-8711.1999.02226.x.

Avgoustidis, A. et al. Constraints on cosmic opacity and beyond the standard model physics from cosmological distance measurements. *Journal of Cosmology and Astroparticle Physics*, 2010. doi: 10.1088/1475-7516/2010/10/024.

Bacon, D., Refregier, A. and Ellis, R. Detection of Weak Gravitational Lensing by Large-scale Structure. *Monthly Notices of the Royal Astronomical Society*, 318(2):625–640, 3 2000. doi: 10.1046/j.1365-8711.2000.03851.x. URL https://arxiv.org/abs/astro-ph/0003008v2.

Bacon, D.J. et al. Numerical simulations of weak lensing measurements. *Monthly Notices of the Royal Astronomical Society*, 325(3):1065–1074, 8 2001. ISSN 0035-8711. doi: 10.1046/J.1365-8711.2001.04507.X. URL https://academic.oup.com/mnras/article/325/3/1065/960747.

Bartelmann, M. and Maturi, M. Weak gravitational lensing. *Scholarpedia*, 12(1):32440, 12 2016. URL http://arxiv.org/abs/1612.06535.

Bartelmann, M. and Schneider, P. Weak Gravitational Lensing. *Physics Report*, 340(4-5):291–472, 12 1999. doi: 10.1016/S0370-1573(00)00082-X. URL http://arxiv.org/abs/astro-ph/9912508http://dx.doi.org/10.1016/S0370-1573(00)00082-X.

Bashkansky, M. and Reintjes, J. Scattering: Stimulated Scattering. In *Encyclopedia of Modern Optics, Five-Volume Set*, pages 330–340. Elsevier Inc., 1 2004. ISBN 9780123693952. doi: 10.1016/B0-12-369395-0/00748-X.

Bass, S.D. and Krzysiak, J. Vacuum energy with mass generation and Higgs bosons. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 803:135351, 4 2020. ISSN 03702693. doi: 10.1016/j.physletb.2020.135351.

Baum. *Problems of extragalactic research* . IAU Symposium, 1962. URL https://cir.nii.ac.jp/crid/1570572700182521984.

Baum. Multicolour Photometry of Spiral Arms. *Spectral Classification and Multicolour Photometry, Proceedings from Symposium no. 24*, 1966. URL https://ui.adsabs.harvard.edu/abs/1966IAUS...24..288B/abstract.

Bautista, J.E. et al. Measurement of baryon acoustic oscillation correlations at z = 2.3 with SDSS DR12 Ly α-Forests. *Astronomy and Astrophysics*, 603:12, 7 2017. ISSN 14320746. doi: 10.1051/0004-6361/201730533. URL http://www.sdss.org/dr12/algorithms/ancillary/boss/.

B.Carter. *Relativistic Fluid Dynamics*. Noto, 1984.

Bekenstein, J.D. Relativistic gravitation theory for the modified Newtonian dynamics paradigm. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 70(8): 083509, 10 2004. ISSN 15502368. doi: 10.1103/PhysRevD.70.083509. URL https://journals.aps.org/prd/abstract/10.1103/PhysRevD.70.083509.

Bellhouse, D.R. The Reverend Thomas Bayes, FRS: A Biography to Celebrate the Tercentenary of His Birth. *https://doi.org/10.1214/088342304000000189*, 19(1):3–43, 2 2004. ISSN 0883-4237. doi: 10.1214/088342304000000189. URL https://projecteuclid.org/journals/statistical-science/volume-19/issue-1/The-Reverend-Thomas-Bayes-FRS--A-Biography-to-Celebrate/10.1214/088342304000000189.fullhttps://projecteuclid.org/journals/statistical-science/volume-19/issue-1/The-Reverend-Thomas-Bayes-FRS--A-Biography-to-Celebrate/10.1214/088342304000000189.short.

Bellman, R. Dynamic programming and stochastic control processes. *Information and Control*, 1(3):228–239, 9 1958. ISSN 0019-9958. doi: 10.1016/S0019-9958(58)80003-0.

Benitez, N. Bayesian Photometric Redshift Estimation. *The Astrophysical Journal*, 536(2): 571–583, 6 2000. ISSN 0004-637X. doi: 10.1086/308947. URL https://ui.adsabs.harvard.edu/abs/2000ApJ...536..571B/abstract.

Benjamin. *Cosmological Probes of Light Relics*. PhD thesis, University of Cambridge, 2018.

Benjamin, J. et al. CFHTLenS tomographic weak lensing: quantifying accurate redshift distributions Downloaded from. *MNRAS*, 431:1547–1564, 2013. doi: 10.1093/mnras/stt276. URL http://mnras.oxfordjournals.org/.

Bernabei, R. et al. First model independent results from DAMA/LIBRA-phase2. *Nuclear Physics and Atomic Energy*, 19(4):307–325, 5 2018. doi: 10.15407/jnpae2018.04.307. URL http://arxiv.org/abs/1805.10486http://dx.doi.org/10.15407/jnpae2018.04.307.

Bernardeau, F., van Waerbeke, L. and Mellier, Y. Weak Lensing Statistics as a Probe of Omega and Power Spectrum. *Astronomy and Astrophysics*, 322(1):1–18, 9 1996. URL https://arxiv.org/abs/astro-ph/9609122v2.

Bernstein, G.M. Shape measurement biases from underfitting and ellipticity gradients. *Monthly Notices of the Royal Astronomical Society*, 406(4):2793–2804, 1 2010. doi: 10.1111/j.1365-2966.2010.16883.x. URL https://arxiv.org/abs/1001.2333v2.

Bessell, M.S. Standard Photometric Systems. *Annu.Rev.Astron.Astrophys*, 43:293–336, 8 2005. ISSN 00664146. doi: 10.1146/ANNUREV.ASTRO.41.082801.100251. URL https://www.annualreviews.org/doi/abs/10.1146/annurev.astro.41.082801.100251.

Bharadwaj, S., Bhavsar, S. and Sheth, J.V. The Size of the Longest Filaments in the Universe. *The Astrophysical Journal*, 606(1):25–31, 11 2003. doi: 10.1086/382140. URL http://arxiv.org/abs/astro-ph/0311342http://dx.doi.org/10.1086/382140.

Bilicki, M. et al. Photometric redshifts for the Kilo-Degree Survey: Machine-learning analysis with artificial neural networks. *Astronomy and Astrophysics*, 616:69, 8 2018. ISSN 14320746. doi: 10.1051/0004-6361/201731942. URL https://doi.org/10.1051/0004-6361/201731942.

Birkedal, A. et al. Little Higgs Dark Matter. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 74(3), 3 2006. doi: 10.1103/PhysRevD.74.035002. URL http://arxiv.org/abs/hep-ph/0603077http://dx.doi.org/10.1103/PhysRevD.74.035002.

Birkinshaw, M., Gull, S.F. and Hardebeck, H. The Sunyaev–Zeldovich effect towards three clusters of galaxies. *Nature 1984 309:5963*, 309(5963):34–35, 1984. ISSN 1476-4687. doi: 10.1038/309034a0. URL https://www.nature.com/articles/309034a0.

Blas, D., Lesgourgues, J. and Tram, T. The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *Journal of Cosmology and Astroparticle Physics*, 2011(7):034, 2011. ISSN 14757516. doi: 10.1088/1475-7516/2011/07/034. URL https://ui.adsabs.harvard.edu/abs/2011JCAP...07..034B/abstract.

Blum, K. DAMA vs. the annually modulated muon background. *ArXiv*, 10 2011. URL http://arxiv.org/abs/1110.0857.

Bolton, A.S. et al. SPECTRAL CLASSIFICATION AND REDSHIFT MEASUREMENT FOR THE SDSS-III BARYON OSCILLATION SPECTROSCOPIC SURVEY. *The Astronomical Journal*, 144(5):144, 10 2012. ISSN 1538-3881. doi: 10.1088/0004-6256/144/5/144. URL https://iopscience.iop.org/article/10.1088/0004-6256/144/5/144https://iopscience.iop.org/article/10.1088/0004-6256/144/5/144/meta.

Bolzonella, M., Miralles, J.M. and Pelló, R. Photometric Redshifts based on standard SED fitting procedures. *Astronomy and Astrophysics*, 363(2):476–492, 3 2000. ISSN 00046361. doi: 10.48550/arxiv.astro-ph/0003380. URL https://arxiv.org/abs/astro-ph/0003380v2.

Bonamente, M. et al. Determination of the Cosmic Distance Scale from Sunyaev-Zel'dovich Effect and Chandra X-Ray Measurements of High-Redshift Galaxy Clusters. *The Astrophysical Journal*, 647(1):25, 8 2006. ISSN 0004-637X. doi: 10.1086/505291. URL https://iopscience.iop.org/article/10.1086/505291https://iopscience.iop.org/article/10.1086/505291/meta.

Bonnett, C. et al. Redshift distributions of galaxies in the Dark Energy Survey Science Verification shear catalogue and implications for weak lensing. *Physical Review D*, 94 (4):042005, 8 2016. ISSN 24700029. doi: 10.1103/PHYSREVD.94.042005/FIGURES/1/SMALL. URL https://ui.adsabs.harvard.edu/abs/2016PhRvD..94d2005B/abstract.

Bonnett, C. Using neural networks to estimate redshift distributions. An application to CFHTLenS. *Monthly Notices of the Royal Astronomical Society*, 449(1):1043–1056, 5 2015. ISSN 0035-8711. doi: 10.1093/MNRAS/STV230. URL https://academic.oup.com/mnras/article/449/1/1043/1302387.

Bordag, M., Mohideen, U. and Mostepanenko, V.M. New Developments in the Casimir Effect. *Physics Report*, 353(1-3):1–205, 6 2001. doi: 10.1016/S0370-1573(01)00015-1. URL http://arxiv.org/abs/quant-ph/0106045http://dx.doi.org/10.1016/S0370-1573(01)00015-1.

Bordoloi, R., Lilly, S.J. and Amara, A. Photo-z performance for precision cosmology. *Monthly Notices of the Royal Astronomical Society*, 406(2):881–895, 8 2010. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2010.16765.X. URL https://academic.oup.com/mnras/article/406/2/881/999103.

Borgani, S. and Guzzo, L. X-ray clusters of galaxies as tracers of structure in the Universe. *Nature 2001 409:6816*, 409(6816):39–45, 1 2001. ISSN 1476-4687. doi: 10.1038/35051000. URL https://www.nature.com/articles/35051000.

Born, M. and Wolf, E. Principles of Optics: 60th Anniversary Edition. *Principles of Optics*, 12 2019. doi: 10.1017/9781108769914. URL https://www.cambridge.org/core/books/principles-of-optics/9D54D6FF0317074912CB285C3FF7341C.

Boruah, S.S., Hudson, M.J. and Lavaux, G. Cosmic flows in the nearby Universe: new peculiar velocities from SNe and cosmological constraints. *Monthly Notices of the Royal Astronomical Society*, 498(2):2703–2718, 12 2019. doi: 10.1093/mnras/staa2485. URL http://arxiv.org/abs/1912.09383http://dx.doi.org/10.1093/mnras/staa2485.

Boylan-Kolchin, M., Bullock, J.S. and Kaplinghat, M. Too big to fail? The puzzling darkness of massive Milky Way subhaloes. *Monthly Notices of the Royal Astronomical Society: Letters*, 415(1):1–6, 2 2011a. doi: 10.1111/j.1745-3933.2011.01074.x. URL http://arxiv.org/abs/1103.0007http://dx.doi.org/10.1111/j.1745-3933.2011.01074.x.

Boylan-Kolchin, M., Bullock, J.S. and Kaplinghat, M. The Milky Way's bright satellites as an apparent failure of LCDM. *Monthly Notices of the Royal Astronomical Society*, 422(2):1203–1218, 11 2011b. doi: 10.1111/j.1365-2966.2012.20695.x. URL http://arxiv.org/abs/1111.2048http://dx.doi.org/10.1111/j.1365-2966.2012.20695.x.

Boynton, P.E. Objects at the Highest Redshift. *Symposium - International Astronomical Union*, 92:293–303, 1980. ISSN 0074-1809. doi: 10.1017/S0074180900068686. URL https://www.cambridge.org/core/journals/symposium-international-astronomical-union/article/objects-at-the-highest-redshift/4B2030396D8C6C9AD8E19680995E29FB.

Brammer, G.B., van Dokkum, P.G. and Coppi, P. EAZY: A Fast, Public Photometric Redshift Code. *The Astrophysical Journal*, 686(2):1503–1513, 7 2008. doi: 10.1086/591786. URL http://arxiv.org/abs/0807.1533http://dx.doi.org/10.1086/591786.

Bridle, S. and King, L. Dark energy constraints from cosmic shear power spectra: impact of intrinsic alignments on photometric redshift requirements. *New Journal of Physics*, 9, 5 2007. doi: 10.1088/1367-2630/9/12/444. URL http://arxiv.org/abs/0705.0166http://dx.doi.org/10.1088/1367-2630/9/12/444.

Bridle, S. et al. Results of the GREAT08 Challenge: An image analysis competition for cosmological lensing. *Monthly Notices of the Royal Astronomical Society*, 405(3):2044–2061, 8 2009. doi: 10.1111/j.1365-2966.2010.16598.x. URL https://arxiv.org/abs/0908.0945v1.

Briel, U.G. and Henry, J.P. An X-ray Temperature Map of Coma. *ArXiv*, 11 1997. URL http://arxiv.org/abs/astro-ph/9711237.

Brinckmann, T. and Lesgourgues, J. MontePython 3: boosted MCMC sampler and other features. *Physics of the Dark Universe*, 24, 4 2018. ISSN 22126864. doi: 10.48550/arxiv.1804.07261. URL https://arxiv.org/abs/1804.07261v2.

Broadhurst, T.J. et al. Large-scale distribution of galaxies at the Galactic poles. *Nature*, 343(6260):726–728, 1990. ISSN 00280836. doi: 10.1038/343726a0. URL https://www.nature.com/articles/343726a0.

Brodwin, M. et al. The Canada-France Deep Fields Survey. III. Photometric Redshift Distribution to I AB = 24 . *The Astrophysical Journal Supplement Series*, 162(1):20–37, 1 2006. ISSN 0067-0049. doi: 10.1086/497990/FULLTEXT/. URL https://iopscience.iop.org/article/10.1086/497990https://iopscience.iop.org/article/10.1086/497990/meta.

Brown, A.G. et al. Gaia Data Release 2 - Summary of the contents and survey properties. *Astronomy & Astrophysics*, 616:A1, 8 2018. ISSN 0004-6361. doi: 10.1051/0004-6361/201833051. URL https://www.aanda.org/articles/aa/full_html/2018/08/aa33051-18/aa33051-18.htmlhttps://www.aanda.org/articles/aa/abs/2018/08/aa33051-18/aa33051-18.html.

Brown, M.L. et al. Measurement of intrinsic alignments in galaxy ellipticities. *Monthly Notices of the Royal Astronomical Society*, 333(3):501–509, 9 2000. doi: 10.1046/j.1365-8711.2002.05354.x. URL http://arxiv.org/abs/astro-ph/0009499http://dx.doi.org/10.1046/j.1365-8711.2002.05354.x.

Brown, M.L. and Battye, R.A. Polarization as an indicator of intrinsic alignment in radio weak lensing. *Monthly Notices of the Royal Astronomical Society*, 410(3):2057–2074, 5 2010. doi: 10.1111/j.1365-2966.2010.17583.x. URL http://arxiv.org/abs/1005.1926http://dx.doi.org/10.1111/j.1365-2966.2010.17583.x.

Bruzual A., G. et al. On the interpretation of colors of faint galaxies. *ApJ*, 241:25–40, 10 1980. ISSN 0004-637X. doi: 10.1086/158314. URL https://ui.adsabs.harvard.edu/abs/1980ApJ...241...25B/abstract.

Bryan, G.L. and Norman, M.L. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. *The Astrophysical Journal*, 495(1):80–99, 3 1998. ISSN 0004-637X. doi: 10.1086/305262/FULLTEXT/. URL https://iopscience.iop.org/article/10.1086/305262https://iopscience.iop.org/article/10.1086/305262/meta.

Bucher, M. Physics of the cosmic microwave background anisotropy. *International Journal of Modern Physics D*, 24(2), 1 2015. doi: 10.1142/S0218271815300049. URL http://arxiv.org/abs/1501.04288http://dx.doi.org/10.1142/S0218271815300049.

Bull, P. et al. Beyond ΛCDM: Problems, solutions, and the road ahead. *Physics of the Dark Universe*, 12:56–99, 12 2015. doi: 10.1016/j.dark.2016.02.001. URL http://arxiv.org/abs/1512.05356http://dx.doi.org/10.1016/j.dark.2016.02.001.

Burns, J. et al. Global 21-cm Cosmology from the Farside of the Moon. *ArXiv*, 3 2021. doi: 10.48550/arxiv.2103.05085. URL https://arxiv.org/abs/2103.05085v1.

Cabayol, L. et al. The PAU survey: Estimating galaxy photometry with deep learning. *Monthly Notices of the Royal Astronomical Society*, 506(3):4048–4069, 4 2021. doi: 10.1093/mnras/stab1909. URL http://arxiv.org/abs/2104.02778http://dx.doi.org/10.1093/mnras/stab1909.

Camelio, G. and Lombardi, M. On the origin of intrinsic alignment in cosmic shear measurements: an analytic argument. *Astronomy & Astrophysics*, 575: A113, 3 2015. ISSN 0004-6361. doi: 10.1051/0004-6361/201425016. URL https://www.aanda.org/articles/aa/full_html/2015/03/aa25016-14/aa25016-14.htmlhttps://www.aanda.org/articles/aa/abs/2015/03/aa25016-14/aa25016-14.html.

Cao, S. and Zhu, Z.H. Cosmic equation of state from combined angular diameter distances: Does the tension with luminosity distances exist? *Physical Review*, 90(8), 10 2014. ISSN 15502368. doi: 10.1103/PHYSREVD.90.083006.

Cao, S. et al. Ultra-compact structure in intermediate-luminosity radio quasars: Building a sample of standard cosmological rulers and improving the dark energy constraints up to z ˜ 3. *Astronomy and Astrophysics*, 606:15, 10 2017. ISSN 14320746. doi: 10.1051/0004-6361/201730551. URL http://nrl.northumbria.ac.uk/.

Capozziello, S., Carloni, S. and Troisi, A. Quintessence without scalar fields. *Recent Res.Dev.Astron.Astrophys*, 1(625), 3 2003. URL http://arxiv.org/abs/astro-ph/0303041.

Capozziello, S., Alberto Mantica, C. and Guido Molinari, L. COSMOLOGICAL PERFECT-FLUIDS IN f (R) GRAVITY. *International Journal of Geometric Methods in Modern Physics*, 16(1), 2018.

Capranico, F., Merkel, P.M. and Schäfer, B.M. Intrinsic ellipticity correlations of galaxies: models, likelihoods and interplay with weak lensing. *Monthly Notices of the Royal Astronomical Society*, 435(1):194–206, 10 2013. ISSN 0035-8711. doi: 10.1093/MNRAS/STT1269. URL https://academic.oup.com/mnras/article/435/1/194/1104047.

Carliles, S. et al. RANDOM FORESTS FOR PHOTOMETRIC REDSHIFTS. *The Astrophysical Journal*, 712(1):511, 3 2010. ISSN 0004-637X. doi: 10.1088/0004-637X/712/1/511. URL https://iopscience.iop.org/article/10.1088/0004-637X/712/1/511https://iopscience.iop.org/article/10.1088/0004-637X/712/1/511/meta.

Carlsten, S.G. et al. Wavelength Dependent PSFs and their impact on Weak Lensing Measurements. *MNRAS*, 000(0000):0–000, 2018. URL https://hsc-release.mtk.nao.ac.jp/.

Carroll, S.M. The Cosmological Constant. *Living Reviews in Relativity*, 4(1), 4 2000. doi: 10.12942/lrr-2001-1. URL http://arxiv.org/abs/astro-ph/0004075http://dx.doi.org/10.12942/lrr-2001-1.

Casimir. On the attraction between two perfectly conducting plates. *Indag. Math.*, 10(4): 261–263, 1948.

Catalan, P. and Theuns, T. Evolution of the angular momentum of protogalaxies from tidal torques: Zel'dovich approximation. *Monthly Notices of the Royal Astronomical Society*, 282(2):436–454, 9 1996. ISSN 0035-8711. doi: 10.1093/MNRAS/282.2.436. URL https://academic.oup.com/mnras/article/282/2/436/1040262.

Catinella, B., Giovanelli, R. and Haynes, M.P. Template Rotation Curves for Disk Galaxies. *The Astrophysical Journal*, 640(2):751–761, 4 2006. ISSN 0004-637X. doi: 10.1086/500171. URL https://iopscience.iop.org/article/10.1086/500171https://iopscience.iop.org/article/10.1086/500171/meta.

Cembranos, J.A.R., Dobado, A. and Maroto, A.L. Brane-world dark matter. *Physical Review Letters*, 90(24):4, 2 2003. doi: 10.1103/PhysRevLett.90.241301. URL http://arxiv.org/abs/hep-ph/0302041http://dx.doi.org/10.1103/PhysRevLett.90.241301.

Chabanier, S. et al. The one-dimensional power spectrum from the SDSS DR14 Lyα forests. *Journal of Cosmology and Astroparticle Physics*, 2019(7), 12 2018. doi: 10.1088/1475-7516/2019/07/017. URL http://arxiv.org/abs/1812.03554http://dx.doi.org/10.1088/1475-7516/2019/07/017.

Chabanier, S., Millea, M. and Palanque-Delabrouille, N. Matter power spectrum: from Lyα forest to CMB scales. *Monthly Notices of the Royal Astronomical Society*, 489(2):2247–2253, 5 2019. doi: 10.1093/mnras/stz2310. URL http://arxiv.org/abs/1905.08103http://dx.doi.org/10.1093/mnras/stz2310.

Chang, C. et al. Atmospheric point spread function interpolation for weak lensing in short exposure imaging data. *Monthly Notices of the Royal Astronomical Society*, 427(3): 2572–2587, 12 2012. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2012.22134.X. URL https://academic.oup.com/mnras/article/427/3/2572/1106605.

Chiosi, C., Bertelli, G. and Bressan, A. No Title. *A&A*, 196(84), 1988.

Choi, A. et al. CFHTLenS and RCSLenS: testing photometric redshift distributions using angular cross-correlations with spectroscopic galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 463(4):3737–3754, 12 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw2241. URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw2241.

Choi, K. et al. Search for Neutrinos from Annihilation of Captured Low-Mass Dark Matter Particles in the Sun by Super-Kamiokande. *Physical Review Letters*, 114(14):141301, 4 2015. ISSN 10797114. doi: 10.1103/PhysRevLett.114.141301. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.114.141301.

Cholis, I., Linden, T. and Hooper, D. Anti-Deuterons and Anti-Helium Nuclei from Annihilating Dark Matter. *Physical Review D*, 102(10), 1 2020. doi: 10.1103/PhysRevD.102.103019. URL http://arxiv.org/abs/2001.08749http://dx.doi.org/10.1103/PhysRevD.102.103019.

Clarkson, C. Establishing homogeneity of the universe in the shadow of dark energy, 7 2012. ISSN 16310705.

Clausius, R. XVI. On a mechanical theorem applicable to heat . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 40(265):122–127, 8 1870. ISSN 1941-5982. doi: 10.1080/14786447008640370. URL https://www.tandfonline.com/doi/abs/10.1080/14786447008640370.

Clowe, D. et al. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, 8 2006. doi: 10.1086/508162. URL http://arxiv.org/abs/astro-ph/0608407http://dx.doi.org/10.1086/508162.

Coil, A.L. The large-scale structure of the universe. *Planets, Stars and Stellar Systems: Volume 6: Extragalactic Astronomy and Cosmology*, pages 387–421, 1 2013. doi: 10.1007/978-94-007-5609-0{_}8/COVER. URL https://link.springer.com/referenceworkentry/10.1007/978-94-007-5609-0_8.

Coil, A.L. et al. THE PRISM MULTI-OBJECT SURVEY (PRIMUS). I. SURVEY OVERVIEW AND CHARACTERISTICS. *The Astrophysical Journal*, 741(1):8, 10 2011. ISSN 0004-637X. doi: 10.1088/0004-637X/741/1/8. URL https://iopscience.iop.org/article/10.1088/0004-637X/741/1/8https://iopscience.iop.org/article/10.1088/0004-637X/741/1/8/meta.

Cole, S. et al. The 2dF galaxy redshift survey: Near-infrared galaxy luminosity functions. *Monthly Notices of the Royal Astronomical Society*, 326(1):255–273, 9 2001. ISSN 00358711. doi: 10.1046/J.1365-8711.2001.04591.X/2/M{_}326-1-255-FIG010.JPEG. URL https://academic.oup.com/mnras/article/326/1/255/1026734.

Coleman, G.D., Wu, C.C. and Weedman, D.W. Colors and magnitudes predicted for high redshift galaxies. *The Astrophysical Journal Supplement Series*, 43:393, 7 1980. ISSN 0067-0049. doi: 10.1086/190674. URL http://adsabs.harvard.edu/doi/10.1086/190674.

Collaboration, P. et al. Planck 2018 results. VI. Cosmological parameters. *Astronomy and Astrophysics*, 641, 7 2018. doi: 10.1051/0004-6361/201833910. URL https://arxiv.org/abs/1807.06209v3.

Comer, G.L., Peter, P. and Andersson, N. Multi-fluid cosmology: An illustration of fundamental principles. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 85(10), 11 2011. doi: 10.1103/PhysRevD.85.103006. URL http://arxiv.org/abs/1111.5043http://dx.doi.org/10.1103/PhysRevD.85.103006.

Connolly, A.J. et al. Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry. *The Astronomical Journal*, 110:2655, 12 1995. ISSN 00046256. doi: 10.1086/117720. URL https://www.researchgate.net/publication/1810942_Slicing_Through_Multicolor_Space_Galaxy_Redshifts_From_Broadband_Photometry.

Coulton, W.R. et al. Exploring the brighter fatter effect with the Hyper Suprime-Cam. *The Astronomical Journal*, 155(6):258, 11 2017. doi: 10.3847/1538-3881/aac08d. URL https://arxiv.org/abs/1711.06273v1.

Cruz, M. et al. Detection of a non-Gaussian Spot in WMAP. *Monthly Notices of the Royal Astronomical Society*, 356(1):29–40, 5 2004. doi: 10.1111/j.1365-2966.2004.08419.x/abs/. URL https://arxiv.org/abs/astro-ph/0405341v2.

Cuesta-Lazaro, C. et al. Gravitational corrections to light propagation in a perturbed FLRW-universe and corresponding weak lensing spectra. *Monthly Notices of the Royal Astronomical Society*, 477(1):741–754, 1 2018. doi: 10.1093/mnras/sty672. URL https://arxiv.org/abs/1801.03325v1.

Cypriano, E.S. et al. Cosmic shear requirements on the wavelength dependence of telescope point spread functions. *Monthly Notices of the Royal Astronomical Society*, 405(1): 494–502, 6 2010. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2010.16461.X. URL https://academic.oup.com/mnras/article/405/1/494/1023433.

D'Agostini, G. Fits, and especially linear fits, with errors on both axes, extra variance of the data points and other complications. *ArXiv*, 11 2005. doi: 10.48550/arxiv.physics/0511182. URL https://arxiv.org/abs/physics/0511182v1.

Dahlen, T. et al. A CRITICAL ASSESSMENT OF PHOTOMETRIC REDSHIFT METHODS: A CANDELS INVESTIGATION. *The Astrophysical Journal*, 775(2):93, 9 2013. ISSN 0004-637X. doi: 10.1088/0004-637X/775/2/93. URL https://iopscience.iop.org/article/10.1088/0004-637X/775/2/93https://iopscience.iop.org/article/10.1088/0004-637X/775/2/93/meta.

Dai, X. et al. Mid-infrared galaxy luminosity functions from the agn and galaxy evolution survey. *Astrophysical Journal*, 697(1):506–521, 2009. ISSN 15384357. doi: 10.1088/0004-637X/697/1/506. URL https://ui.adsabs.harvard.edu/abs/2009ApJ...697..506D/abstract.

Das, A. et al. Perfect fluid cosmological Universes: One equation of state and the most general solution. *Pramana*, 90:19, 2017. doi: 10.1007/s12043-017-1511-z. URL https://doi.org/10.1007/s12043-017-1511-z.

Davis, T.M., Lineweaver, C.H. and Webb, J.K. Solutions to the tethered galaxy problem in an expanding universe and the observation of receding blueshifted objects. *American Journal of Physics*, 71(358), 2003.

Davis, T.M. et al. The effect of peculiar velocities on supernova cosmology. *Astrophysical Journal*, 741(1), 12 2010. doi: 10.1088/0004-637X/741/1/67. URL http://arxiv.org/abs/1012.2912http://dx.doi.org/10.1088/0004-637X/741/1/67.

de Bernardis, P. et al. A Flat Universe from High-Resolution Maps of the Cosmic Microwave Background Radiation. *Nature*, 404(6781):955–959, 4 2000. doi: 10.1038/35010035. URL http://arxiv.org/abs/astro-ph/0004404http://dx.doi.org/10.1038/35010035.

de Jong, J.T.A. et al. The Kilo-Degree Survey. *Experimental Astronomy, Volume 35, Issue 1-2, pp. 25-44*, 35:25–44, 6 2012. ISSN 0922-6435. doi: 10.1007/s10686-012-9306-1. URL http://arxiv.org/abs/1206.1254http://dx.doi.org/10.1007/s10686-012-9306-1.

de Jong, J.T.A. et al. The first and second data releases of the Kilo-Degree Survey. *Astronomy & Astrophysics*, 582:A62, 10 2015. ISSN 0004-6361. doi: 10.1051/0004-6361/201526601. URL http://www.aanda.org/10.1051/0004-6361/201526601.

De Jong, J.T. et al. The third data release of the Kilo-Degree Survey and associated data products. *Astronomy & Astrophysics*, 604:A134, 8 2017. ISSN 0004-6361. doi: 10.1051/0004-6361/201730747. URL https://www.aanda.org/articles/aa/full_html/2017/08/aa30747-17/aa30747-17.htmlhttps://www.aanda.org/articles/aa/abs/2017/08/aa30747-17/aa30747-17.html.

De Paolis, F. et al. MACHOs as brown dwarfs. *ArXiv*, 3 1998. URL http://arxiv.org/abs/astro-ph/9803200.

de Vaucouleurs, G., Vaucouleurs, d. and Gerard. Classification and Morphology of External Galaxies. *HDP*, 53:275, 1959. doi: 10.1007/978-3-642-45932-0{_}7. URL https://ui.adsabs.harvard.edu/abs/1959HDP....53..275D/abstract.

Desjacques, V., Jeong, D. and Schmidt, F. Large-Scale Galaxy Bias. *Physics Reports*, 733: 1–193, 11 2016. doi: 10.1016/j.physrep.2017.12.002. URL http://arxiv.org/abs/1611.09787http://dx.doi.org/10.1016/j.physrep.2017.12.002.

Díaz Tello, J. et al. Star formation activity in Balmer break galaxies at z< 1.5. *Astronomy & Astrophysics*, 587:A136, 3 2016. ISSN 0004-6361. doi: 10.1051/0004-6361/201526315. URL https://www.aanda.org/articles/aa/full_html/2016/03/aa26315-15/aa26315-15.htmlhttps://www.aanda.org/articles/aa/abs/2016/03/aa26315-15/aa26315-15.html.

D'Isanto, A. and Polsterer, K.L. Photometric redshift estimation via deep learning. *Astronomy and Astrophysics*, 609, 6 2017. doi: 10.1051/0004-6361/201731326. URL http://arxiv.org/abs/1706.02467http://dx.doi.org/10.1051/0004-6361/201731326.

Donato, F., Fornengo, N. and Salati, P. Antideuterons as a signature of supersymmetric dark matter. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 62 (4):1–13, 8 2000. ISSN 05562821. doi: 10.1103/PhysRevD.62.043003. URL https://journals.aps.org/prd/abstract/10.1103/PhysRevD.62.043003.

Dong, T., An, D. and H. Kim, N. Prognostics 102: Efficient Bayesian-Based Prognostics Algorithm in MATLAB. *Fault Detection, Diagnosis and Prognosis*, 2 2020. doi: 10.5772/INTECHOPEN.82781.

Doroshkevich, A.G. et al. Large-scale galaxy distribution in the Las Campanas Redshift Survey. *Monthly Notices of the Royal Astronomical Society*, 322(2):369–388, 4 2001. ISSN 00358711. doi: 10.1046/j.1365-8711.2001.04121.x. URL https://academic.oup.com/mnras/article/322/2/369/963406.

Dressler, A. et al. Spectroscopy and photometry of elliptical galaxies. I - A new distance estimator. *The Astrophysical Journal*, 313:42, 2 1987. ISSN 0004-637X. doi: 10.1086/164947. URL https://ui.adsabs.harvard.edu/abs/1987ApJ...313...42D/abstract.

Driver, S. Galaxy Spectra, 2014. URL http://star-www.st-and.ac.uk/~spd3/Teaching/PHYS1002/phys1002_lecture6.pdf.

Durrer, R. The Cosmic Microwave Background: The history of its experimental investigation and its significance for cosmology. *Classical and Quantum Gravity*, 32(12), 6 2015. doi: 10.1088/0264-9381/32/12/124007. URL http://arxiv.org/abs/1506.01907http://dx.doi.org/10.1088/0264-9381/32/12/124007.

Dvali, G., Gabadadze, G. and Porrati, M. 4D Gravity on a Brane in 5D Minkowski Space. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 485 (1-3):208–214, 5 2000. doi: 10.1016/S0370-2693(00)00669-9. URL http://arxiv.org/abs/hep-th/0005016http://dx.doi.org/10.1016/S0370-2693(00)00669-9.

Dyson, F.W. et al. A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *RSPTA*, 220 (571-581):291–333, 1 1920. ISSN 1364-503X. doi: 10.1098/RSTA.1920.0009. URL https://ui.adsabs.harvard.edu/abs/1920RSPTA.220..291D/abstract.

Eddington, A.S. Einstein shift and Doppler shift, 1 1926. ISSN 00280836. URL https://www.nature.com/articles/117086a0.

Edge, A. et al. The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift - NASA/ADS. *The Messenger*, 154:32–34, 12 2013. URL https://ui.adsabs.harvard.edu/abs/2013Msngr.154...32E/abstract.

Efstathiou, G., Bond, J.R. and White, S.D. COBE background radiation anisotropics and large-scale structure in the Universe. *Monthly Notices of the Royal Astronomical Society*, 258(1):1P–6P, 9 1992. ISSN 13652966. doi: 10.1093/mnras/258.1.1P. URL https://academic.oup.com/mnras/article/258/1/1P/967069.

Efstathiou, G. To H0 or not to H0? *Monthly Notices of the Royal Astronomical Society*, 505(3):3866–3872, 6 2021. ISSN 0035-8711. doi: 10.1093/MNRAS/STAB1588. URL https://academic.oup.com/mnras/article/505/3/3866/6293858.

Einstein, A. Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes. *Annalen der Physik*, 340(10):898–908, 1 1911. ISSN 1521-3889. doi: 10.1002/ANDP.19113401005. URL https://onlinelibrary.wiley.com/doi/full/10.1002/andp.19113401005https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19113401005https://onlinelibrary.wiley.com/doi/10.1002/andp.19113401005.

Einstein, A. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354(7):769–822, 1 1916. ISSN 1521-3889. doi: 10.1002/ANDP.19163540702. URL https://onlinelibrary.wiley.com/doi/full/10.1002/andp.19163540702https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19163540702https://onlinelibrary.wiley.com/doi/10.1002/andp.19163540702.

Einstein, A. and Rosen, N. On gravitational waves. *Journal of the Franklin Institute*, 223(1): 43–54, 1 1937. ISSN 00160032. doi: 10.1016/S0016-0032(37)90583-0.

Einstein, A., Einstein and A. Erklarung der Perihelionbewegung der Merkur aus der allgemeinen Relativitatstheorie. *SPAW*, 47:831–839, 1915. URL https://ui.adsabs.harvard.edu/abs/1915SPAW...47..831E/abstract.

Einstein, A. On the Electrodynamics of Moving Bodies. *Annalen der Physik*, 1905.

Einstein, A. Cosmological Considerations in the General Theory of Relativity. *Sitzungsber.Preuss.Akad.Wiss.Berlin (Math.Phys.)*, 1917:142–152, 1917.

Einstein, A. and Cohen, E.R. The Meaning of Relativity. *Physics Today*, 9(10):30, 1 1956. ISSN 0031-9228. doi: 10.1063/1.3059795. URL https://physicstoday.scitation.org/doi/abs/10.1063/1.3059795.

Eisenstein, D.J. et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *The Astrophysical Journal*, 633(2):560–574, 1 2005. doi: 10.1086/466512. URL http://arxiv.org/abs/astro-ph/0501171http://dx.doi.org/10.1086/466512.

Eisenstein, D.J. and Weinberg, D.H.e.a. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems. *Astronomical Journal*, 142(3):19, 1 2011. doi: 10.1088/0004-6256/142/3/72. URL http://arxiv.org/abs/1101.1529http://dx.doi.org/10.1088/0004-6256/142/3/72.

Ellis, J. et al. Gravitino Dark Matter in the CMSSM. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 588(1-2):7–16, 12 2003. doi: 10.1016/j.physletb.2004.03.021. URL http://arxiv.org/abs/hep-ph/0312262http://dx.doi.org/10.1016/j.physletb.2004.03.021.

Er, X. et al. Calibration of colour gradient bias in shear measurement using HST/CANDELS data. *Monthly Notices of the Royal Astronomical Society*, 476(4):5645–5657, 8 2017. doi: 10.1093/mnras/sty685. URL https://arxiv.org/abs/1708.06085v3.

Erdogdu, P. et al. Reconstructed Density and Velocity Fields from the 2MASS Redshift Survey. *Monthly Notices of the Royal Astronomical Society*, 373(1):45–64, 9 2006. doi: 10.1111/j.1365-2966.2006.11049.x. URL http://arxiv.org/abs/astro-ph/0610005http://dx.doi.org/10.1111/j.1365-2966.2006.11049.x.

Etherington, I. LX. On the definition of distance in general relativity . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 15(100):761–773, 4 1933. ISSN 1941-5982. doi: 10.1080/14786443309462220. URL https://www.tandfonline.com/doi/abs/10.1080/14786443309462220.

Euclid Collaboration, E. et al. Euclid Preparation IV. Impact of undetected galaxies on weak lensing shear measurements. *Astronomy & Astrophysics*, 627:A59, 1 2019. URL http://arxiv.org/abs/1902.00044.

Fairall, A.P., Fairall and P., A. A caution to those who measure galaxy redshifts. *The Observatory*, 112:286–287, 1992. ISSN 0029-7704. URL https://ui.adsabs.harvard.edu/abs/1992Obs...112..286F/abstract.

Famaey, B. and McGaugh, S. Modified Newtonian Dynamics (MOND): Observational Phenomenology and Relativistic Extensions. *Living Reviews in Relativity*, 15, 12 2011. doi: 10.12942/lrr-2012-10. URL http://arxiv.org/abs/1112.3960http://dx.doi.org/10.12942/lrr-2012-10.

Feldman, H.A., Kaiser, N. and Peacock, J.A. Power Spectrum Analysis of Three-Dimensional Redshift Surveys. *The Astrophysical Journal*, 426:23, 4 1993. doi: 10.1086/174036. URL http://arxiv.org/abs/astro-ph/9304022http://dx.doi.org/10.1086/174036.

Feldmann, R. et al. The Zurich Extragalactic Bayesian Redshift Analyzer (ZEBRA) and its first application: COSMOS. *Monthly Notices of the Royal Astronomical Society*, 372 (2):565–577, 9 2006. doi: 10.1111/j.1365-2966.2006.10930.x. URL http://arxiv.org/abs/astro-ph/0609044http://dx.doi.org/10.1111/j.1365-2966.2006.10930.x.

Feng, J.L. Dark Matter Candidates from Particle Physics and Methods of Detection. *Annual Review of Astronomy and Astrophysics*, 48:495–545, 3 2010. doi: 10.1146/annurev-astro-082708-101659. URL http://arxiv.org/abs/1003.0904http://dx.doi.org/10.1146/annurev-astro-082708-101659.

Feng, J.L., Rajaraman, A. and Takayama, F. Superweakly Interacting Massive Particles. *Physical Review Letters*, 91(1), 2 2003. doi: 10.1103/PhysRevLett.91.011302. URL http://arxiv.org/abs/hep-ph/0302215http://dx.doi.org/10.1103/PhysRevLett.91.011302.

Ferreras, I. et al. Confronting MOND and TeVeS with strong gravitational lensing over galactic scales: An extended survey. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 86(8):083507, 10 2012. ISSN 15507998. doi: 10.1103/PhysRevD.86.083507. URL https://journals.aps.org/prd/abstract/10.1103/PhysRevD.86.083507.

Fisher, K.B. et al. The power spectrum of IRAS galaxies. *The Astrophysical Journal*, 402:42, 1 1993. ISSN 0004-637X. doi: 10.1086/172110. URL https://ui.adsabs.harvard.edu/abs/1993ApJ...402...42F/abstract.

Forman, W. et al. The fourth Uhuru catalog of X-ray sources. *ApJS*, 38:357–412, 12 1978. ISSN 0067-0049. doi: 10.1086/190561. URL https://ui.adsabs.harvard.edu/abs/1978ApJS..38..357F/abstract.

Francis, M.J. et al. Expanding Space: the Root of all Evil? *Publications of the Astronomical Society of Australia*, 24(2):95–102, 7 2007. doi: 10.1071/AS07019. URL http://arxiv.org/abs/0707.0380http://dx.doi.org/10.1071/AS07019.

Freedman, W.L. et al. Calibration of the Tip of the Red Giant Branch (TRGB). *arXiv*, 2 2020. doi: 10.3847/1538-4357/ab7339. URL http://arxiv.org/abs/2002.01550http://dx.doi.org/10.3847/1538-4357/ab7339.

Freese, K. Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars. *EAS Publications Series*, 36:113–126, 12 2008. doi: 10.1051/eas/0936016. URL http://arxiv.org/abs/0812.4005http://dx.doi.org/10.1051/eas/0936016.

Friedman, A. Über die Krümmung des Raumes. *Zeitschrift für Physik*, 10(1):377–386, 12 1922. ISSN 14346001. doi: 10.1007/BF01332580.

Friedmann, A. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, 21(1):326–332, 12 1924. ISSN 14346001. doi: 10.1007/BF01328280. URL https://ui.adsabs.harvard.edu/abs/1924ZPhy...21..326F/abstract.

Fry, J.N. and Gaztanaga, E. Biasing and Hierarchical Statistics in Large-scale Structure. *The Astrophysical Journal*, 413:447, 2 1993. doi: 10.1086/173015. URL http://arxiv.org/abs/astro-ph/9302009http://dx.doi.org/10.1086/173015.

Gamow, G. The evolution of the universe. *Nature*, 162(4122):680–682, 1948. ISSN 00280836. doi: 10.1038/162680a0. URL https://www.nature.com/articles/162680a0.

Gelmini, G.B. TASI 2014 Lectures: The Hunt for Dark Matter. *ArXiv*, 2 2015. URL http://arxiv.org/abs/1502.01320.

Giblin, J.T., Mertens, J.B. and Starkman, G.D. Observable Deviations from Homogeneity in an Inhomogeneous Universe. *The Astrophysical Journal*, 833(2):247, 8 2016. doi: 10.3847/1538-4357/833/2/247. URL http://arxiv.org/abs/1608.04403http://dx.doi.org/10.3847/1538-4357/833/2/247.

Gomes, Z. et al. Improving photometric redshift estimation using GPZ: Size information, post processing, and improved photometry. *Monthly Notices of the Royal Astronomical Society*, 475(1):331–342, 3 2018. ISSN 13652966. doi: 10.1093/MNRAS/STX3187.

Graff, P. et al. SkyNet: an efficient and robust neural network training tool for machine learning in astronomy. *Monthly Notices of the Royal Astronomical Society*, 441(2):1741–1759, 6 2014. ISSN 0035-8711. doi: 10.1093/MNRAS/STU642. URL https://academic.oup.com/mnras/article/441/2/1741/1071156.

Gromov, A., Baryshev, Y. and Teerikorpi, P. Two-fluid matter-quintessence FLRW models: energy transfer and the equation of state of the universe. *Astronomy and Astrophysics*, 415(3):813–820, 9 2002. doi: 10.1051/0004-6361:20031693. URL http://arxiv.org/abs/astro-ph/0209458http://dx.doi.org/10.1051/0004-6361:20031693.

Group, P.D. et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8):1–2093, 8 2020. doi: 10.1093/PTEP/PTAA104. URL https://academic.oup.com/ptep/article/2020/8/083C01/5891211.

Guidorzi, C. et al. Improved constraints on H0 from a combined analysis of gravitational-wave and electromagnetic emission from GW170817, 10 2017. ISSN 23318422. URL https://doi.org/10.3847/2041-8213/aaa009.

Gunn, J.E. et al. Investigations of the optical fields of 3CR radio sources to faint limiting magnitudes – IV. *Monthly Notices of the Royal Astronomical Society*, 194(1):111–123, 1 1981. ISSN 0035-8711. doi: 10.1093/MNRAS/194.1.111. URL https://academic.oup.com/mnras/article/194/1/111/985910.

Gunn, J.E. and Peterson, B.A. On the Density of Neutral Hydrogen in Intergalactic Space. *The Astrophysical Journal*, 142:1633, 11 1965. ISSN 0004-637X. doi: 10.1086/148444. URL https://ui.adsabs.harvard.edu/abs/1965ApJ...142.1633G/abstract.

Ha, T. et al. Classification of the FRW universe with a cosmological constant and a perfect fluid of the equation of state p = wρ. *General Relativity and Gravitation*, 44:1433–1458, 2012.

Hamilton, A.J.S. Uncorrelated modes of the non-linear power spectrum. *Monthly Notices of the Royal Astronomical Society*, 312(2):257–284, 2 2000. ISSN 0035-8711. doi: 10.1046/J.1365-8711.2000.03071.X. URL https://academic.oup.com/mnras/article/312/2/257/972743.

Hamilton, A.J.S., Gott, J. Richard, I. and Weinberg, D. The topology of the large-scale structure of the universe. *The Astrophysical Journal*, 309:1, 10 1986. ISSN 0004-637X. doi: 10.1086/164571. URL https://ui.adsabs.harvard.edu/abs/1986ApJ...309....1H/abstract.

Hand, N. et al. First measurement of the cross-correlation of CMB lensing and galaxy lensing. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 91(6), 3 2015. ISSN 15502368. doi: 10.1103/PHYSREVD.91.062001.

Hankin, R.K.S. A Generalization of the Dirichlet Distribution. *Journal of statistical software*, 33(11), 2 2010.

Harnois-Déraps, J. et al. Baryons, neutrinos, feedback and weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 450(2):1212–1223, 4 2015. ISSN 13652966. doi: 10.1093/mnras/stv646. URL https://ui.adsabs.harvard.edu/abs/2015MNRAS.450.1212H/abstract.

Harvey, D. et al. The non-gravitational interactions of dark matter in colliding galaxy clusters. *Science*, 347(6229):1462–1465, 3 2015. doi: 10.1126/science.1261381. URL http://arxiv.org/abs/1503.07675http://dx.doi.org/10.1126/science.1261381.

Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 4 1970. ISSN 0006-3444. doi: 10.1093/BIOMET/57.1.97. URL https://academic.oup.com/biomet/article/57/1/97/284580.

Heavens, A. Weak Gravitational Lensing. *Lecture Notes in Physics*, 665:585–600, 2008. doi: 10.1007/978-3-540-44767-2{_}18. URL https://link.springer.com/chapter/10.1007/978-3-540-44767-2_18.

Heavens, A. 3D weak lensing. *Monthly Notices of the Royal Astronomical Society*, 343(4): 1327–1334, 4 2003. doi: 10.1046/j.1365-8711.2003.06780.x. URL http://arxiv.org/abs/astro-ph/0304151http://dx.doi.org/10.1046/j.1365-8711.2003.06780.x.

Heavens, A., Alsing, J. and Jaffe, A. Combining Size and Shape in Weak Lensing. *Monthly Notices of the Royal Astronomical Society: Letters*, 433(1):0–000, 2 2013. doi: 10.1093/mnrasl/slt045. URL https://arxiv.org/abs/1302.1584v2.

Heitmann, K. et al. The coyote universe. II. cosmological models and precision emulation of the nonlinear matter power spectrum. *Astrophysical Journal*, 705(1):156–174, 2009. ISSN 15384357. doi: 10.1088/0004-637X/705/1/156. URL https://ui.adsabs.harvard.edu/abs/2009ApJ...705..156H/abstract.

Heros, C.P.d.l. Status of direct and indirect dark matter searches. *Proceedings of Science*, 364, 1 2020. URL http://arxiv.org/abs/2001.06193.

Heymans, C. and Heavens, A. Weak gravitational lensing: Reducing the contamination by intrinsic alignments. *Monthly Notices of the Royal Astronomical Society*, 339(3):711–720, 3 2003. ISSN 00358711. doi: 10.1046/J.1365-8711.2003.06213.X/2/339-3-711-FIG006.JPEG. URL https://academic.oup.com/mnras/article/339/3/711/971367.

Hikage, C. et al. Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data. *Publications of the Astronomical Society of Japan*, 71(2), 4 2019. ISSN 0004-6264. doi: 10.1093/pasj/psz010. URL https://academic.oup.com/pasj/article/doi/10.1093/pasj/psz010/5370019.

Hilbert, S. et al. Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. *Astronomy & Astrophysics*, 499(1):31–43, 5 2009. ISSN 0004-6361. doi: 10.1051/0004-6361/200811054. URL https://www.aanda.org/articles/aa/abs/2009/19/aa11054-08/aa11054-08.html.

Hildebrandt, H., Van Waerbeke, L. and Erben, T. Astrophysics CARS: The CFHTLS-Archive-Research Survey III. First detection of cosmic magnification in samples of normal high-z galaxies. *A&A*, 507:683–691, 2009. doi: 10.1051/0004-6361/200912655.

Hildebrandt, H. et al. PHAT: PHoto-z Accuracy Testing. *Astronomy & Astrophysics*, 523(2):A31, 11 2010. ISSN 0004-6361. doi: 10.1051/0004-6361/201014885. URL https://www.aanda.org/articles/aa/full_html/2010/15/aa14885-10/aa14885-10.htmlhttps://www.aanda.org/articles/aa/abs/2010/15/aa14885-10/aa14885-10.html.

Hildebrandt, H. et al. CFHTLenS: Improving the quality of photometric redshifts with precision photometry. *Monthly Notices of the Royal Astronomical Society*, 421(3):2355–2367, 4 2012. ISSN 00358711. doi: 10.1111/J.1365-2966.2012.20468.X/2/MNRAS0421-2355-F10.JPEG. URL https://academic.oup.com/mnras/article/421/3/2355/1078193.

Hildebrandt, H. et al. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 465(2):1454–1498, 2 2017. ISSN 0035-8711. doi: 10.1093/MNRAS/STW2805. URL https://academic.oup.com/mnras/article/465/2/1454/2417034.

Hildebrandt, H. et al. KiDS+VIKING-450: Cosmic shear tomography with optical+infrared data. *Astronomy and Astrophysics*, 633, 12 2018. doi: 10.1051/0004-6361/201834878. URL http://arxiv.org/abs/1812.06076http://dx.doi.org/10.1051/0004-6361/201834878.

Hildebrandt, H. et al. KiDS-1000 catalogue: Redshift distributions and their calibration. *Astronomy and Astrophysics*, 647, 7 2020. ISSN 14320746. doi: 10.1051/0004-6361/202039018. URL https://arxiv.org/abs/2007.15635v2.

Hirata, C.M. and Seljak, U. Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 70(6):11, 2004. ISSN 15502368. doi: 10.1103/PhysRevD.70.063526. URL https://ui.adsabs.harvard.edu/abs/2004PhRvD..70f3526H/abstract.

Hivon, E. et al. MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets. *The Astrophysical Journal*, 567(1):2–17, 3 2002. ISSN 0004-637X. doi: 10.1086/338126/FULLTEXT/. URL https://iopscience.iop.org/article/10.1086/338126https://iopscience.iop.org/article/10.1086/338126/meta.

Hobson, M.P., Efstathiou, G.P. and Lasenby, A.N. *General Relativity: An Introduction for Physicists*. Cambridge University Press, 2 2006. ISBN 9780521829519. doi: 10.1017/CBO9780511790904. URL https://www.cambridge.org/core/books/general-relativity/ADB781FA6C986E3C941B3ECB5D1EF01A.

Hogg, D.W. Distance measures in cosmology. *ArXiv*, 5 1999. URL http://arxiv.org/abs/astro-ph/9905116.

Hogg, D.W. Data analysis recipes: Probability calculus for inference. *ArXiv*, 5 2012. doi: 10.48550/arxiv.1205.4446. URL https://arxiv.org/abs/1205.4446v1.

Hogg, D.W. et al. The K correction. *ArXiv*, 10 2002. doi: 10.48550/arxiv.astro-ph/0210394. URL https://arxiv.org/abs/astro-ph/0210394v1.

Holanda, R.F.L., Lima, J.A.S. and Ribeiro, M.B. Cosmic distance duality relation and the shape of galaxy clusters. *A&A*, 528:14, 2011. doi: 10.1051/0004-6361/201015547.

Hosseini, R. and Bethge, M. Spectral Stacking: Unbiased Shear Estimation for Weak Gravitational Lensing Spectral Stacking: Unbiased Shear Estimation for Weak Gravitational Lensing Spectral Stacking: Unbiased Shear Estimation for Weak Gravitational Lensing. Technical report, Max Planck Institute for Biological Cybernetics, 2009.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 9 1933. ISSN 00220663. doi: 10.1037/H0071325. URL /record/1934-00645-001.

Howley, T. et al. The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data. *Knowledge-Based Systems*, 19(5):363–370, 9 2006. ISSN 0950-7051. doi: 10.1016/J.KNOSYS.2005.11.014.

Hoyle, B. et al. Dark Energy Survey Year 1 Results: redshift distributions of the weak-lensing source galaxies. *Monthly Notices of the Royal Astronomical Society*, 478(1):592–610, 7 2018. ISSN 0035-8711. doi: 10.1093/MNRAS/STY957. URL https://academic.oup.com/mnras/article/478/1/592/4975790.

Hu, W. Power Spectrum Tomography with Weak Lensing. *The Astrophysical Journal*, 522 (1):L21–L24, 9 1999. ISSN 0004637X. doi: 10.1086/312210.

Hu, W. and Dodelson, S. Cosmic Microwave Background Anisotropies. *Annu. Rev. Astron. and Astrophys*, 2002.

Hubble, E. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 3 1929. ISSN 0027-8424. doi: 10.1073/pnas.15.3.168. URL https://www.pnas.org/content/15/3/168https://www.pnas.org/content/15/3/168.abstract.

Huff, E.M. et al. Cosmic shear without shape noise. *ArXiv*, 11 2013. URL https://arxiv.org/abs/1311.1489v2.

Huterer, D. et al. Systematic errors in future weak-lensing surveys: Requirements and prospects for self-calibration. *Monthly Notices of the Royal Astronomical Society*, 366(1):101–114, 2 2006. ISSN 13652966. doi: 10.1111/J.1365-2966.2005.09782.X/3/366-1-101-FIG008.JPEG. URL https://academic.oup.com/mnras/article/366/1/101/1058840.

Ilbert, O. et al. Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *Astronomy & Astrophysics*, 457(3):841–856, 10 2006. ISSN 0004-6361. doi: 10.1051/0004-6361:20065138. URL https://www.aanda.org/articles/aa/abs/2006/39/aa5138-06/aa5138-06.html.

Ilbert, O. et al. COSMOS PHOTOMETRIC REDSHIFTS WITH 30-BANDS FOR 2-deg2. *The Astrophysical Journal*, 690(2):1236, 12 2008. ISSN 0004-637X. doi: 10.1088/0004-637X/690/2/1236. URL https://iopscience.iop.org/article/10.1088/0004-637X/690/2/1236https://iopscience.iop.org/article/10.1088/0004-637X/690/2/1236/meta.

J.A Peacock et al. *Modern Cosmology*. IOP Publishing, Bristol, 2001.

Jaffe, R.L. Casimir effect and the quantum vacuum. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 72(2):1–5, 7 2005. ISSN 15507998. doi: 10.1103/PhysRevD.72.021301. URL https://journals.aps.org/prd/abstract/10.1103/PhysRevD.72.021301.

Jain, B. and Seljak, U. Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes. *The Astrophysical Journal*, 484(2):560–573, 8 1997. ISSN 0004-637X. doi: 10.1086/304372/FULLTEXT/. URL https://iopscience.iop.org/article/10.1086/304372https://iopscience.iop.org/article/10.1086/304372/meta.

Jarvis, M. et al. The DES Science Verification weak lensing shear catalogues. *MNRAS*, 460(2):2245–2281, 8 2016. ISSN 0035-8711. doi: 10.1093/MNRAS/STW990. URL https://ui.adsabs.harvard.edu/abs/2016MNRAS.460.2245J/abstract.

Jeffreys, H. *Theory of probability*. Oxford University Press, 11 1998. ISBN 9780198503682.

Joachimi, B. and Schneider, P. The removal of shear-ellipticity correlations from the cosmic shear signal: Influence of photometric redshift errors on the nulling technique. *Astronomy and Astrophysics*, 507(1):105–129, 5 2009. doi: 10.1051/0004-6361/200912420. URL http://arxiv.org/abs/0905.0393http://dx.doi.org/10.1051/0004-6361/200912420.

Jollife, I.T. and Cadima, J. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 4 2016. ISSN 1364503X. doi: 10.1098/RSTA.2015.0202. URL http://dx.doi.org/10.1098/rsta.2015.0202.

Joudaki, S. et al. KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear. *Astronomy and Astrophysics*, 638:L1, 6 2020. ISSN 14320746. doi: 10.1051/0004-6361/201936154. URL https://ui.adsabs.harvard.edu/abs/2020A&A...638L...1J/abstract.

Kacprzak, T. et al. Sérsic galaxy models in weak lensing shape measurement: model bias, noise bias and their interaction. *MNRAS*, 441:2528–2538, 2014. doi: 10.1093/mnras/stu588. URL https://academic.oup.com/mnras/article/441/3/2528/1108396.

Kaiser, N. and Peacock, J.A. Power-spectrum analysis of one-dimensional redshift surveys. *The Astrophysical Journal*, 379:482, 10 1991. ISSN 0004-637X. doi: 10.1086/170523. URL https://ui.adsabs.harvard.edu/abs/1991ApJ...379..482K/abstract.

Kaiser, N. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 7 1987. ISSN 0035-8711. doi: 10.1093/mnras/227.1.1. URL https://academic.oup.com/mnras/article/227/1/1/1065830.

Kaiser, N., Kaiser and Nick. Weak Gravitational Lensing of Distant Galaxies. *ApJ*, 388:272, 4 1992. ISSN 0004-637X. doi: 10.1086/171151. URL https://ui.adsabs.harvard.edu/abs/1992ApJ...388..272K/abstract.

Kaiser, N. et al. Mapping the Dark Matter with Weak Gravitational Lensing. *ApJ*, 404:441, 2 1993. ISSN 0004-637X. doi: 10.1086/172297. URL https://ui.adsabs.harvard.edu/abs/1993ApJ...404..441K/abstract.

Kaiser, N., Kaiser and Nick. A New Shear Estimator for Weak-Lensing Observations. *ApJ*, 537(2):555–577, 7 2000. ISSN 0004-637X. doi: 10.1086/309041. URL https://ui.adsabs.harvard.edu/abs/2000ApJ...537..555K/abstract.

Kannawadi, A. et al. Towards emulating cosmic shear data: revisiting the calibration of the shear measurements for the Kilo-Degree Survey. *Astronomy & Astrophysics*, 624:A92, 4 2019. ISSN 14320746. doi: 10.1051/0004-6361/201834819. URL https://hdl.handle.net/1887/83079https://scholarlypublications.universiteitleiden.nl/handle/1887/83079/.

Karachentsev, I.D. et al. Peculiar Velocities of 3000 Spiral Galaxies from the 2MFGC Catalog. *Astrophysics*, 49(4):450–461, 11 2006. doi: 10.1007/s10511-006-0044-9. URL http://arxiv.org/abs/astro-ph/0611945http://dx.doi.org/10.1007/s10511-006-0044-9.

Kelly, P.L. et al. Weighing the Giants II: Improved Calibration of Photometry from Stellar Colors and Accurate Photometric Redshifts. *Mon. Not. R. Astron. Soc*, 000:1–21, 2012. URL http://smoka.nao.ac.jp/.

Kiessling, A. et al. Galaxy alignments: Theory, modelling and simulations. *Space Science Reviews*, 193(1-4):67–136, 4 2015. doi: 10.1007/s11214-015-0203-6. URL http://arxiv.org/abs/1504.05546http://dx.doi.org/10.1007/s11214-015-0203-6.

Kilbinger, M. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8), 11 2014. doi: 10.1088/0034-4885/78/8/086901. URL http://arxiv.org/abs/1411.0115http://dx.doi.org/10.1088/0034-4885/78/8/086901.

Kilbinger, M. et al. CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 430(3):2200–2220, 4 2013. ISSN 0035-8711. doi: 10.1093/MNRAS/STT041. URL https://academic.oup.com/mnras/article/430/3/2200/981767.

Kim, A. Beta Distribution — Intuition, Examples, and Derivation | by Aerin Kim | Towards Data Science, 2020. URL https://towardsdatascience.com/beta-distribution-intuition-examples-and-derivation-cf00f4db57af.

Kind, M.C. and Brunner, R.J. TPZ : Photometric redshift PDFs and ancillary information by using prediction trees and random forests. *Monthly Notices of the Royal Astronomical Society*, 432(2):1483–1501, 3 2013a. doi: 10.1093/mnras/stt574. URL http://arxiv.org/abs/1303.7269http://dx.doi.org/10.1093/mnras/stt574.

Kind, M.C. and Brunner, R.J. SOMz: photometric redshift PDFs with self organizing maps and random atlas. *Monthly Notices of the Royal Astronomical Society*, 438(4):3409–3421, 12 2013b. doi: 10.1093/mnras/stt2456. URL http://arxiv.org/abs/1312.5753http://dx.doi.org/10.1093/mnras/stt2456.

Kind, M.C. and Brunner, R.J. Sparse representation of photometric redshift probability density functions: preparing for petascale astronomy. *Monthly Notices of the Royal Astronomical Society*, 441(4):3550–3561, 7 2014. ISSN 0035-8711. doi: 10.1093/MNRAS/STU827. URL https://academic.oup.com/mnras/article/441/4/3550/1229381.

King, L.J. and Schneider, P. Separating cosmic shear from intrinsic galaxy alignments: Correlation function tomography. *Astronomy & Astrophysics*, 398(1):23–30, 1 2003. ISSN 0004-6361. doi: 10.1051/0004-6361:20021614. URL https://www.aanda.org/articles/aa/abs/2003/04/aa3114/aa3114.html.

Kinney, A.L. et al. Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and Their Application to K-Corrections. *ApJ*, 467:38, 8 1996. ISSN 0004-637X. doi: 10.1086/177583. URL https://ui.adsabs.harvard.edu/abs/1996ApJ...467...38K/abstract.

Kippenhahn, R., Weigert, A. and Weiss, A. *Stellar Structure and Evolution*. Astronomy and Astrophysics Library. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-30255-8. doi: 10.1007/978-3-642-30304-3. URL http://link.springer.com/10.1007/978-3-642-30304-3.

Kitching, T.D. et al. 3D cosmic shear: cosmology from CFHTLenS. *Monthly Notices of the Royal Astronomical Society*, 442(2):1326–1349, 8 2014. ISSN 0035-8711. doi: 10.1093/MNRAS/STU934. URL https://academic.oup.com/mnras/article/442/2/1326/985112.

Kitching, T. et al. GRAVITATIONAL LENSING ACCURACY TESTING 2010 (GREAT10) CHALLENGE HANDBOOK. *The Annals of Applied Statistics*, 5(3):2231–2263, 2011. doi: 10.1214/11-AOAS484. URL http://www.pascal-network.org/.

Kitching, T.D. et al. The limits of cosmic shear. *Monthly Notices of the Royal Astronomical Society*, 469(3):2737–2749, 8 2017. ISSN 0035-8711. doi: 10.1093/MNRAS/STX1039. URL https://academic.oup.com/mnras/article/469/3/2737/3786432.

Klypin, A.A. et al. Where are the missing galactic satellites? *The Astrophysical Journal*, 522(1):82–92, 1 1999. doi: 10.1086/307643. URL http://arxiv.org/abs/astro-ph/9901240http://dx.doi.org/10.1086/307643.

Knox, L. and Millea, M. The Hubble Hunter's Guide, 8 2019. ISSN 23318422. URL https://journals.aps.org/prd/abstract/10.1103/PhysRevD.101.043533.

Köhlinger, F. et al. KiDS-450: The tomographic weak lensing power spectrum and constraints on cosmological parameters. *Monthly Notices of the Royal Astronomical Society*, 471(4):4412–4435, 6 2017. doi: 10.1093/mnras/stx1820. URL http://arxiv.org/abs/1706.02892http://dx.doi.org/10.1093/mnras/stx1820.

Kollatschny, W. and Zetzl, M. The shape of broad-line profiles in active galactic nuclei. *Astronomy & Astrophysics*, 549:A100, 1 2013. ISSN 0004-6361. doi: 10.1051/0004-6361/201219411. URL https://www.aanda.org/articles/aa/full_html/2013/01/aa19411-12/aa19411-12.htmlhttps://www.aanda.org/articles/aa/abs/2013/01/aa19411-12/aa19411-12.html.

Koo, D.C., Koo and C., D. Optical multicolors : a poor person's Z machine for galaxies. *AJ*, 90(3):418–440, 3 1985. ISSN 0004-6256. doi: 10.1086/113748. URL https://ui.adsabs.harvard.edu/abs/1985AJ.....90..418K/abstract.

Koo, D.C. Photometric Redshifts: A Perspective from an Old-Timer on Its Past, Present, and Potential. *Photometric Redshifts and High Redshift Galaxies*, page 10, 7 1999. doi: 10.48550/arxiv.astro-ph/9907273. URL https://arxiv.org/abs/astro-ph/9907273v1.

Koposov, S. et al. The Luminosity Function of the Milky Way Satellites. *The Astrophysical Journal*, 686(1):279–291, 6 2007. doi: 10.1086/589911. URL http://arxiv.org/abs/0706.2687http://dx.doi.org/10.1086/589911.

Kraus, H. et al. Eureca - The future of cryogenic dark matter detection in Europe. In *EAS Publications Series*, volume 36, pages 249–255. EDP Sciences, 2009. ISBN 9782759804399. doi: 10.1051/eas/0936035. URL https://www.eas-journal.org/articles/eas/abs/2009/03/eas0936035/eas0936035.html.

Kuijken, K. GaaP: PSF- and aperture-matched photometry using shapelets. *Astronomy & Astrophysics*, 482(3):1053–1067, 5 2008. ISSN 0004-6361. doi: 10.1051/0004-6361:20066601. URL http://www.aanda.org/10.1051/0004-6361:20066601.

Kuijken, K. OmegaCAM: ESO's Newest Imager - NASA/ADS. *The Messenger*, 146:8–11, 12 2011. URL https://ui.adsabs.harvard.edu/abs/2011Msngr.146....8K/abstract.

Kuijken, K. et al. The fourth data release of the Kilo-Degree Survey: Ugri imaging and nine-band optical-IR photometry over 1000 square degrees. *Astronomy and Astrophysics*, 625:A2, 5 2019. ISSN 14320746. doi: 10.1051/0004-6361/201834918. URL https://ui.adsabs.harvard.edu/abs/2019A&A...625A...2K/abstract.

Kuijken, K. et al. Gravitational lensing analysis of the Kilo-Degree Survey. *Monthly Notices of the Royal Astronomical Society*, 454(4):3500–3532, 12 2015. ISSN 0035-8711. doi: 10.1093/MNRAS/STV2140. URL https://academic.oup.com/mnras/article/454/4/3500/993789.

Kunnen, J. *A Search for Dark Matter in the Center of the Earth with the IceCube Neutrino Detector*. PhD thesis, Vrije U., Brussels, 2015.

Lahav, O. Observational Tests for the Cosmological Principle and World Models. *Structure Formation in the Universe*, pages 131–142, 2001. doi: 10.1007/978-94-010-0540-1{_}7/COVER. URL https://link.springer.com/chapter/10.1007/978-94-010-0540-1_7.

Lahav, O. and Liddle, A.R. The Cosmological Parameters 2014. *ArXiv*, 1 2014. URL http://arxiv.org/abs/1401.1389.

Lake, S.E. UCLA UCLA Electronic Theses and Dissertations Title A WISE Measurement of the 2.4 μm Galaxy Luminosity Function and its Implications for the Extragalactic Background Light at 3.4 μm, 2017. URL https://escholarship.org/uc/item/6hh4c2nt.

Lançon, A. and Wood, P.R. A library of 0.5 to 2.5 μm spectra of luminous cool stars. *Astronomy and Astrophysics Supplement Series*, 146(2):217–249, 10 2000. ISSN 0365-0138. doi: 10.1051/AAS:2000269. URL https://aas.aanda.org/articles/aas/abs/2000/17/h2019/h2019.html.

Laureijs, R. et al. Euclid Definition Study Report. *Euclid Consortium*, 10 2011. URL http://arxiv.org/abs/1110.3193.

Lavaux, G. et al. Cosmic flow from two micron all-sky redshift survey: The origin of cosmic microwave background dipole and implications for ΛcDM cosmology. *Astrophysical Journal*, 709(1):483–498, 1 2010. ISSN 15384357. doi: 10.1088/0004-637X/709/1/483. URL https://iopscience.iop.org/article/10.1088/0004-637X/709/1/483https://iopscience.iop.org/article/10.1088/0004-637X/709/1/483/meta.

Le Fèvre, O. et al. The VIMOS VLT deep survey First epoch VVDS-deep survey. *Astronomy & Astrophysics*, 439(3):845–862, 9 2005. ISSN 0004-6361. doi: 10.1051/0004-6361:20041960. URL https://www.aanda.org/articles/aa/abs/2005/33/aa1960-04/aa1960-04.html.

Le Verrier, U.J., Verrier, L. and J., U. Theorie du mouvement de Mercure. Technical report, Annales de l'Observatoire imperial de Paris, 1859. URL https://ui.adsabs.harvard.edu/abs/1859AnPar...5....1L/abstract.

Leauthaud, A. et al. Weak Gravitational Lensing with COSMOS: Galaxy Selection and Shape Measurements. *The Astrophysical Journal Supplement Series*, 172(1):219–238, 2007. URL http://snap.lbl.gov.

Lebach, D.E. et al. Measurement of the Solar Gravitational Deflection of Radio Waves Using Very-Long-Baseline Interferometry. *Physical Review Letters*, 75(8):1439, 8 1995. doi: 10.1103/PhysRevLett.75.1439. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.75.1439.

Leclercq, F. *Bayesian large-scale structure inference and cosmic web analysis*. PhD thesis, Institut d'Astrophysique de Paris, Pris, 9 2015.

Leistedt, B., Mortlock, D.J. and Peiris, H.V. Hierarchical Bayesian inference of galaxy redshift distributions from photometric surveys. *Monthly Notices of the Royal Astronomical Society*, 460(4):4258–4267, 2 2016. doi: 10.1093/mnras/stw1304. URL http://arxiv.org/abs/1602.05960http://dx.doi.org/10.1093/mnras/stw1304.

Lejeune, T., Cuisinier, F. and Buser, R. A standard stellar library for evolutionary synthesis: II. The M dwarf extension. *Astronomy and Astrophysics Supplement Series*, 130(1):65–75, 5 1998. ISSN 03650138. doi: 10.1051/aas:1998405. URL https://ui.adsabs.harvard.edu/abs/1998A&AS..130...65L/abstract.

Lematre, A.G. and Eddington, A.S. The Expanding Universe. *Monthly Notices of the Royal Astronomical Society*, 91(5):490–501, 3 1931. ISSN 0035-8711. doi: 10.1093/mnras/91.5.490. URL https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/91.5.490.

Lewis, A. Galaxy shear estimation from stacked images. *Monthly Notices of the Royal Astronomical Society*, 398(1):471–476, 9 2009. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2009.15161.X. URL https://academic.oup.com/mnras/article/398/1/471/1099856.

Lieu, R. Proving the conservation of surface brightness during the strong and weak lensing of light by an isothermal sphere. *ArXiv*, 9 2004. URL https://arxiv.org/abs/astro-ph/0409655v3.

Lilly, S.J. et al. The zCOSMOS 10k-Bright Spectroscopic Sample. *ApJS*, 184(2):218–229, 2009. ISSN 0067-0049. doi: 10.1088/0067-0049/184/2/218. URL https://ui.adsabs.harvard.edu/abs/2009ApJS..184..218L/abstract.

Lima, M. et al. Estimating the redshift distribution of photometric galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 390(1):118–130, 10 2008. ISSN 00358711. doi: 10.1111/j.1365-2966.2008.13510.x. URL https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2008.13510.x.

Limber, D.N., Limber and Nelson, D. The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *ApJ*, 117:134, 1 1953. ISSN 0004-637X. doi: 10.1086/145672. URL https://ui.adsabs.harvard.edu/abs/1953ApJ...117..134L/abstract.

Liske, J. et al. Galaxy And Mass Assembly (GAMA): end of survey report and data release 2. *Monthly Notices of the Royal Astronomical Society*, 452(2):2087–2126, 9 2015. ISSN 0035-8711. doi: 10.1093/MNRAS/STV1436. URL https://academic.oup.com/mnras/article/452/2/2087/1069711.

Liu, L. and Chronopoulos, C. The Hydrogen 21-cm Line and Its Applications to Radio Astrophysics. *ArXiv*, 2008.

Liu, S. The Dirichlet Distribution: What Is It and Why Is It Useful? | Built In, 2022. URL https://builtin.com/data-science/dirichlet-distribution.

Lo, K.H., Young, K. and Lee, B.Y.P. Advance of perihelion. *American Journal of Physics*, 81(9):695–702, 9 2013. ISSN 0002-9505. doi: 10.1119/1.4813067. URL http://aapt.scitation.org/doi/10.1119/1.4813067.

Loverde, M. and Afshordi, N. Extended Limber approximation. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 78(12):123506, 12 2008. ISSN 15507998. doi: 10.1103/PhysRevD.78.123506. URL https://ui.adsabs.harvard.edu/abs/2008PhRvD..78l3506L/abstract.

LSST Science Collaboration, L.S. et al. LSST Science Book, Version 2.0. *ArXiv*, 12 2009. URL http://arxiv.org/abs/0912.0201.

Lupton, R.H., Gunn, J.E. and Szalay, A.S. A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements. *The Astronomical Journal*, 118(3):1406–1410, 9 1999. ISSN 00046256. doi: 10.1086/301004/FULLTEXT/. URL https://iopscience.iop.org/article/10.1086/301004https://iopscience.iop.org/article/10.1086/301004/meta.

Lusso, E. et al. Tension with the flat LambdaCDM model from a high redshift Hubble Diagram of supernovae, quasars and gamma-ray bursts. *Astronomy and Astrophysics*, 628, 7 2019. doi: 10.1051/0004-6361/201936223. URL https://arxiv.org/abs/1907.07692v4.

Maartens, R. Is the Universe homogeneous? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1957):5115–5137, 4 2011. doi: 10.1098/rsta.2011.0289. URL http://arxiv.org/abs/1104.1300http://dx.doi.org/10.1098/rsta.2011.0289.

Mackay, D.J.C. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 9 1995. ISBN 0521642981. URL http://www.inference.phy.cam.ac.uk/mackay/itila/.

Maggiore, M. et al. Science Case for the Einstein Telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(3), 12 2019. doi: 10.1088/1475-7516/2020/03/050. URL http://arxiv.org/abs/1912.02622http://dx.doi.org/10.1088/1475-7516/2020/03/050.

Malz, A.I. How not to obtain the redshift distribution from probabilistic redshift estimates: Under what conditions is it not inappropriate to estimate the redshift distribution N(z) by stacking photo-z PDFs? *Physical Review D*, 103(8), 1 2021. doi: 10.1103/PhysRevD.103.083502. URL http://arxiv.org/abs/2101.04675http://dx.doi.org/10.1103/PhysRevD.103.083502.

Malz, A.I. and Hogg, D.W. How to obtain the redshift distribution from probabilistic redshift estimates. *arXiv*, 7 2020. URL http://arxiv.org/abs/2007.12178.

Mandelbaum, R. Weak lensing for precision cosmology. *Annual Review of Astronomy and Astrophysics*, 56:393–433, 10 2017. doi: 10.1146/annurev-astro-081817-051928. URL https://arxiv.org/abs/1710.03235v1.

Mandelbaum, R. et al. Detection of large-scale intrinsic ellipticity-density correlation from the Sloan Digital Sky Survey and implications for weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 367(2):611–626, 4 2006. ISSN 00358711. doi: 10.1111/J.1365-2966.2005.09946.X/2/M{_}367-2-611-EQ036.JPEG. URL https://academic.oup.com/mnras/article/367/2/611/1012784.

Maraston, C. Evolutionary synthesis of stellar populations: a modular tool. *Monthly Notices of the Royal Astronomical Society*, 300(3):872–892, 11 1998. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1998.01947.x. URL https://ui.adsabs.harvard.edu/abs/1998MNRAS.300..872M/abstract.

Maraston, C. Evolutionary population synthesis: Models, analysis of the ingredients and application to high-z galaxies. *Monthly Notices of the Royal Astronomical Society*, 362(3):799–825, 9 2005. ISSN 00358711. doi: 10.1111/J.1365-2966.2005.09270.X/2/362-3-799-FIG030.JPEG. URL https://academic.oup.com/mnras/article/362/3/799/976732.

Martin, S.P. A Supersymmetry Primer. *Adv.Ser.Direct.High Energy Phys*, 18:1–98, 9 1997. doi: 10.1142/9789812839657{_}0001. URL http://arxiv.org/abs/hep-ph/9709356http://dx.doi.org/10.1142/9789812839657_0001.

Massarotti, M., Iovino, A. and Buzzoni, A. A critical appraisal of the SED fitting method to estimate photometric redshifts. *Astronomy & Astrophysics*, 368(1):74–85, 3 2001. ISSN 0004-6361. doi: 10.1051/0004-6361:20000553. URL https://www.aanda.org/articles/aa/abs/2001/10/aah2454/aah2454.html.

Massey, R. et al. Origins of weak lensing systematics, and requirements on future instrumentation (or knowledge of instrumentation). *Mon. Not. R. Astron. Soc*, 000:1–19, 2011.

Masters, K.L. et al. SFI++ I : A New I -Band Tully-Fisher Template, the Cluster Peculiar Velocity Dispersion, and H 0 . *The Astrophysical Journal*, 653(2):861–880, 12 2006. ISSN 0004-637X. doi: 10.1086/508924. URL https://iopscience.iop.org/article/10.1086/508924https://iopscience.iop.org/article/10.1086/508924/meta.

Masters, K.L., Springob, C.M. and Huchra, J.P. 2MTF. I. the tully-fisher relation in the two micron all sky survey J, H, and K bands. *Astronomical Journal*, 135 (5):1738–1748, 5 2008. ISSN 00046256. doi: 10.1088/0004-6256/135/5/1738. URL https://iopscience.iop.org/article/10.1088/0004-6256/135/5/1738https://iopscience.iop.org/article/10.1088/0004-6256/135/5/1738/meta.

Masters, K.L. et al. Galaxy Zoo: Unwinding the Winding Problem - Observations of Spiral Bulge Prominence and Arm Pitch Angles Suggest Local Spiral Galaxies are Winding. *Monthly Notices of the Royal Astronomical Society*, 487(2):1808–1820, 4 2019. doi: 10.1093/mnras/stz1153. URL http://arxiv.org/abs/1904.11436http://dx.doi.org/10.1093/mnras/stz1153.

Mead, A.J. et al. An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models. *Monthly Notices of the Royal Astronomical Society*, 454(2):1958–1975, 12 2015. ISSN 13652966. doi: 10.1093/mnras/stv2036. URL https://ui.adsabs.harvard.edu/abs/2015MNRAS.454.1958M/abstract.

Melchior, P. et al. Limitations on shapelet-based weak-lensing measurements. *A&A*, 510:75, 2010. doi: 10.1051/0004-6361/200912785.

Menanteau, F. et al. The Atacama Cosmology Telescope: ACT-CL J0102-4915 "El Gordo," a Massive Merging Cluster at Redshift 0.87. *Astrophysical Journal*, 748(1), 9 2011. doi: 10.1088/0004-637x/748/1/7. URL https://arxiv.org/abs/1109.0953v3.

Metcalf, R.B. and Silk, J. New constraints on macroscopic compact objects as dark matter candidates from gravitational lensing of type Ia supernovae. *Physical Review Letters*, 98(7):071302, 2 2007. ISSN 00319007. doi: 10.1103/PhysRevLett.98.071302. URL https://link.aps.org/doi/10.1103/PhysRevLett.98.071302.

Metropolis, N. et al. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 12 1953. ISSN 0021-9606. doi: 10.1063/1.1699114. URL https://aip.scitation.org/doi/abs/10.1063/1.1699114.

Meyers, J.E. and Burchat, P.R. IMPACT OF ATMOSPHERIC CHROMATIC EFFECTS ON WEAK LENSING MEASUREMENTS. *The Astrophysical Journal*, 807(2):182, 7 2015. ISSN 0004-637X. doi: 10.1088/0004-637X/807/2/182. URL https://iopscience.iop.org/article/10.1088/0004-637X/807/2/182https://iopscience.iop.org/article/10.1088/0004-637X/807/2/182/meta.

Mihalas, D. and Routly, P.M. *Galactic astronomy*. Princeton University Press, 9 1968. ISBN 9780716703266.

Milgrom, M. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *The Astrophysical Journal*, 270:365, 7 1983. ISSN 0004-637X. doi: 10.1086/161130.

Monaco, P. Approximate methods for the generation of dark matter halo catalogs in the age of precision cosmology. *Galaxies*, 4(4), 5 2016. URL http://arxiv.org/abs/1605.07752.

Moore, B. et al. Dark Matter Substructure in Galactic Halos. *The Astrophysical Journal*, 524(1):L19–L22, 7 1999. doi: 10.1086/312287. URL http://arxiv.org/abs/astro-ph/9907411http://dx.doi.org/10.1086/312287.

Moore, C.C. Ergodic theorem, ergodic theory, and statistical mechanics, 2 2015. ISSN 10916490. URL www.pnas.org/cgi/doi/10.1073/pnas.1421798112.

Mootoovaloo, A. et al. Parameter Inference for Weak Lensing using Gaussian Processes and MOPED. *MNRAS*, 000:1–15, 2020.

Moster, B.P. et al. A cosmic variance cookbook. *Astrophysical Journal*, 731(2):113, 4 2011. ISSN 15384357. doi: 10.1088/0004-637X/731/2/113. URL http://astro.berkeley.edu/.

Moulin, E. Astroparticle Physics with H.E.S.S.: recents results and nearfuture prospects. *EPJ Web of Conferences*, 209:01054, 2019. doi: 10.1051/epjconf/201920901054. URL https://doi.org/10.1051/e.

Mueller, M.W. et al. Propagating star formation and irregular structure in spiral galaxies. *ApJ*, 210:670–678, 12 1976. ISSN 0004-637X. doi: 10.1086/154873. URL https://ui.adsabs.harvard.edu/abs/1976ApJ...210..670M/abstract.

Munoz, R.R. et al. Exploring Halo Substructure with Giant Stars: The Dynamics and Metallicity of the Dwarf Spheroidal in Bootes. *The Astrophysical Journal*, 650(1):L51–L54, 6 2006. doi: 10.1086/508685. URL http://arxiv.org/abs/astro-ph/0606271http://dx.doi.org/10.1086/508685.

Murase, K. and Shoemaker, I.M. Detecting Asymmetric Dark Matter in the Sun with Neutrinos. *Physical Review D*, 94(6), 6 2016. doi: 10.1103/PhysRevD.94.063512. URL http://arxiv.org/abs/1606.03087http://dx.doi.org/10.1103/PhysRevD.94.063512.

Nadler, E.O. et al. Milky Way Satellite Census. II. Galaxy–Halo Connection Constraints Including the Impact of the Large Magellanic Cloud. *arXiv*, 19:31, 12 2019. doi: 10.3847/1538-4357/ab846a. URL http://arxiv.org/abs/1912.03303http://dx.doi.org/10.3847/1538-4357/ab846a.

Nagamine, K. and Loeb, A. Future Evolution of Nearby Large-Scale Structure in a Universe Dominated by a Cosmological Constant. *New Astronomy*, 8(5):439–448, 4 2002. doi: 10.1016/S1384-1076(02)00234-8. URL http://arxiv.org/abs/astro-ph/0204249http://dx.doi.org/10.1016/S1384-1076(02)00234-8.

Narlikar, J.V. Spectral shifts in general relativity. *American Journal of Physics*, 62(10): 903–907, 10 1994. ISSN 0002-9505. doi: 10.1119/1.17679. URL http://aapt.scitation.org/doi/10.1119/1.17679.

NASA & ESA. Hubble Galaxy Classification, 8 1999. URL https://esahubble.org/images/heic9902o/.

Neal, R.M. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 1–592, 6 2012. doi: 10.1201/b10905. URL http://arxiv.org/abs/1206.1901http://dx.doi.org/10.1201/b10905.

Newman, J. et al. Spectroscopic Needs for Imaging Dark Energy Experiments: Photometric Redshift Training and Calibration. *APh*, 63:81–100, 9 2013a. doi: 10.1016/j.astropartphys.2014.06.007. URL http://arxiv.org/abs/1309.5384http://dx.doi.org/10.1016/j.astropartphys.2014.06.007.

Newman, J.A. Calibrating Redshift Distributions beyond Spectroscopic Limits with Cross-Correlations. *The Astrophysical Journal*, 684(1):88–101, 9 2008. ISSN 0004-637X. doi: 10.1086/589982. URL https://ui.adsabs.harvard.edu/abs/2008ApJ...684...88N/abstract.

Newman, J.A. et al. The deep2 galaxy redshift survey: Design, observations, data reduction, and redshifts. *Astrophysical Journal, Supplement Series*, 208(1):5, 9 2013b. ISSN 00670049. doi: 10.1088/0067-0049/208/1/5. URL https://ui.adsabs.harvard.edu/abs/2013ApJS..208....5N/abstract.

Newman, J.A. et al. Spectroscopic needs for imaging dark energy experiments. *Astroparticle Physics*, 63:81–100, 3 2015. ISSN 0927-6505. doi: 10.1016/J.ASTROPARTPHYS.2014.06.007.

Newton, I. *Opticks: or, a treatise of the reflections,refraction,inflections,and colours of light*. Sam. Smith, and Benj. Walford,, 1704.

Nicolaou, C. et al. The Impact of Peculiar Velocities on the Estimation of the Hubble Constant from Gravitational Wave Standard Sirens. *Monthly Notices of the Royal Astronomical Society*, 495(1):90–97, 9 2019. doi: 10.1093/mnras/staa1120. URL http://arxiv.org/abs/1909.09609http://dx.doi.org/10.1093/mnras/staa1120.

Nusser, A. and Dekel, A. Omega and the initial fluctuations from velocity and density fields. *The Astrophysical Journal*, 405:437, 3 1993. ISSN 0004-637X. doi: 10.1086/172376. URL https://ui.adsabs.harvard.edu/abs/1993ApJ...405..437N/abstract.

Oke, J.B. et al. Secondary standard stars for absolute spectrophotometry. *ApJ*, 266:713–717, 3 1983. ISSN 0004-637X. doi: 10.1086/160817. URL https://ui.adsabs.harvard.edu/abs/1983ApJ...266..713O/abstract.

Okumura, T., Jing, Y.P. and Li, C. Intrinsic Ellipticity Correlation of SDSS Luminous Red Galaxies and Misalignment with their Host Dark Matter Halos. *Astrophysical Journal*, 694(1):214–221, 9 2008. doi: 10.1088/0004-637X/694/1/214. URL http://arxiv.org/abs/0809.3790http://dx.doi.org/10.1088/0004-637X/694/1/214.

Okura, Y. and Futamase, T. Elliptical Weighted HOLICs for Weak Lensing Shear Measurement part3:Random Count Noise Effect for Image's Moments in Weak Lensing Analysis. *Astrophysical Journal*, 771(1), 8 2012. doi: 10.1088/0004-637x/771/1/37. URL https://arxiv.org/abs/1208.3564v1.

O'Raifeartaigh, C. et al. One Hundred Years of the Cosmological Constant: from 'Superfluous Stunt' to Dark Energy. *European Physical Journal H*, 43(1):73–117, 11 2017. doi: 10.1140/epjh/e2017-80061-7. URL http://arxiv.org/abs/1711.06890http://dx.doi.org/10.1140/epjh/e2017-80061-7.

Ostriker, J.P. et al. Generation of Microwave Background Fluctuations from Nonlinear Perturbations at the ERA of Galaxy Formation. *ApJL*, 306:L51, 7 1986. ISSN 0004-637X. doi: 10.1086/184704. URL https://ui.adsabs.harvard.edu/abs/1986ApJ...306L..51O/abstract.

Padmanabhan, N. et al. Calibrating photometric redshifts of luminous red galaxies. *Monthly Notices of the Royal Astronomical Society*, 359(1):237–250, 5 2005. ISSN 00358711. doi: 10.1111/J.1365-2966.2005.08915.X/3/359-1-237-FIG012.GIF. URL https://academic.oup.com/mnras/article/359/1/237/983557.

Partridge, R.B., Partridge and B., R. A. Search for Primeval Galaxies at High Redshifts. *ApJ*, 192:241–250, 9 1974. ISSN 0004-637X. doi: 10.1086/153055. URL https://ui.adsabs.harvard.edu/abs/1974ApJ...192..241P/abstract.

Peacock, J.A. *Cosmological Physics*. Cambridge University Press, 12 1998. doi: 10.1017/cbo9780511804533. URL https://www.cambridge.org/core/books/cosmological-physics/3EA0C68658270C9C3CC48E07A035C327.

Peacock, J.A. A diatribe on expanding space. *ArXiv*, 9 2008. URL http://arxiv.org/abs/0809.4573.

Peacock, J.A. and Dodds, S.J. Reconstructing the linear power spectrum of cosmological mass fluctuations. *Monthly Notices of the Royal Astronomical Society*, 267(4):1020–1034, 4 1994. ISSN 0035-8711. doi: 10.1093/MNRAS/267.4.1020. URL https://academic.oup.com/mnras/article/267/4/1020/1227458.

Pearson. LIII. On lines and planes of closest fit to systems of points in space. *https://doi.org/10.1080/14786440109462720*, 2(11):559–572, 11 1901. ISSN 1941-5982. doi: 10.1080/14786440109462720. URL https://www.tandfonline.com/doi/abs/10.1080/14786440109462720.

Peebles, P.J.E. The Void Phenomenon. *The Astrophysical Journal*, 557(2):495–504, 8 2001. ISSN 0004-637X. doi: 10.1086/322254. URL www.astro.princeton.edu/strauss/ors/index.html.

Penzias, A.A. and Wilson, R.W. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419, 7 1965. ISSN 0004-637X. doi: 10.1086/148307. URL https://ui.adsabs.harvard.edu/abs/1965ApJ...142..419P/abstract.

Perivolaropoulos, L. Sub-millimeter Spatial Oscillations of Newton's Constant: Theoretical Models and Laboratory Tests. *Physical Review D*, 95(8), 11 2016. doi: 10.1103/physrevd.95.084050. URL https://arxiv.org/abs/1611.07293v3.

Perivolaropoulos, L. and Skara, F. Challenges for ΛCDM: An update. *New Astronomy Reviews*, 95, 5 2021. URL http://arxiv.org/abs/2105.05208.

Perlmutter, S. et al. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *The Astrophysical Journal*, 517(2):565–586, 12 1998. doi: 10.1086/307221. URL http://arxiv.org/abs/astro-ph/9812133http://dx.doi.org/10.1086/307221.

Petri, A., Haiman, Z. and May, M. On the validity of the Born approximation for beyond-Gaussian weak lensing observables. *ArXiv*, 2017.

Petters, A.O., Levine, H. and Wambsganss, J. *Singularity Theory and Gravitational Lensing*. Birkhäuser Boston, 2001. doi: 10.1007/978-1-4612-0145-8.

Piattella, O.F. Lecture Notes in Cosmology. *arXiv*, 2 2018. doi: 10.1007/978-3-319-95570-4. URL http://arxiv.org/abs/1803.00070http://dx.doi.org/10.1007/978-3-319-95570-4.

Pitrou, C., Uzan, J.P. and Pereira, T.S. Weak lensing B-modes on all scales as a probe of local isotropy. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 87(4), 3 2012. doi: 10.1103/physrevd.87.043003. URL https://arxiv.org/abs/1203.6029v3.

Pitrou, C. et al. Precision big bang nucleosynthesis with improved Helium-4 predictions. *Physics Reports*, 754:1–66, 1 2018. doi: 10.1016/j.physrep.2018.04.005. URL https://arxiv.org/abs/1801.08023v2.

Planelles, S., Schleicher, D.R. and Bykov, A.M. Large-Scale Structure Formation: From the First Non-linear Objects to Massive Galaxy Clusters. *Space Science Reviews 2014 188:1*, 188(1):93–139, 4 2014. ISSN 1572-9672. doi: 10.1007/S11214-014-0045-7. URL https://link.springer.com/article/10.1007/s11214-014-0045-7.

Polchinski, J. The Cosmological Constant and the String Landscape. *The Quantum Structure of Space and Time : Proceedings of the 23rd Solvay Conference on Physics. Brussels, Belgium*, 3 2006. URL http://arxiv.org/abs/hep-th/0603249.

Polletta, M. et al. Spectral Energy Distributions of Hard X-ray selected AGNs in the XMDS Survey. *ApJ*, 663(1):81–102, 3 2007. doi: 10.1086/518113. URL http://arxiv.org/abs/astro-ph/0703255http://dx.doi.org/10.1086/518113.

Polsterer, K.L., D'Isanto, A. and Gieseke, F. Uncertain Photometric Redshifts. *ArXiv*, 8 2016. doi: 10.48550/arxiv.1608.08016. URL https://arxiv.org/abs/1608.08016v1.

Popesso, P. et al. RASS-SDSS galaxy cluster survey. II. A unified picture of the cluster luminosity function. *A&A*, 433(2):415–429, 4 2005. ISSN 0004-6361. doi: 10.1051/0004-6361:20041870. URL https://ui.adsabs.harvard.edu/abs/2005A&A...433..415P/abstract.

Pössel, M. Interpretations of cosmic expansion: anchoring conceptions and misconceptions. *Physics Education*, 55(6), 8 2020. doi: 10.1088/1361-6552/aba3b1. URL http://arxiv.org/abs/2008.07776http://dx.doi.org/10.1088/1361-6552/aba3b1.

Pound, R.V. and Rebka, G.A. Apparent weight of photons. *Physical Review Letters*, 4 (7):337–341, 4 1960. ISSN 00319007. doi: 10.1103/PhysRevLett.4.337. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.4.337.

Pritchard, J. Nested Sampling, 2016. URL https://www.imperial.ac.uk/media/imperial-college/research-centres-and-groups/astrophysics/public/icic/data-analysis-workshop/2016/NestedSampling_JRP.pdf.

Puschell, Owen and Laning. Near-infrared photometry of distant radio galaxies-Spectral flux distributions and redshift estimates. *The Astrophysical Journal*, 257, 1982. URL https://scholar.google.co.uk/citations?view_op=view_citation&hl=en&user=t3Y8bXAAAAAJ&citation_for_view=t3Y8bXAAAAAJ:7PzlFSSx8tAC.

Qi, J.Z. et al. Testing the Etherington's distance duality relation at higher redshifts: the combination of radio quasars and gravitational waves. *ArXiv*, 2019.

Raichoor, A. et al. THE NEXT GENERATION VIRGO CLUSTER SURVEY. XV. THE PHOTOMETRIC REDSHIFT ESTIMATION FOR BACKGROUND SOURCES. *The Astrophysical Journal*, 797(2):102, 12 2014. ISSN 0004-637X. doi: 10.1088/0004-637X/797/2/102. URL https://iopscience.iop.org/article/10.1088/0004-637X/797/2/102https://iopscience.iop.org/article/10.1088/0004-637X/797/2/102/meta.

Randall, L. and Xu, W.L. Searching for dark photon dark matter with cosmic ray antideuterons. *Journal of High Energy Physics 2020 2020:5*, 2020(5):1–22, 5 2020. ISSN 1029-8479. doi: 10.1007/JHEP05(2020)081. URL https://link.springer.com/article/10.1007/JHEP05(2020)081.

Rauch, M. The Lyman Alpha Forest in the Spectra of QSOs. *Annual Review of Astronomy and Astrophysics*, 36(1):267–316, 6 1998. doi: 10.1146/annurev.astro.36.1.267. URL http://arxiv.org/abs/astro-ph/9806286http://dx.doi.org/10.1146/annurev.astro.36.1.267.

Refregier, A. Shapelets: I. A Method for Image Analysis. *Mon. Not. R. Astron. Soc*, 000: 1–13, 2001.

Refregier, A. and Bacon, D. Shapelets: II. A Method for Weak Lensing Measurements. *Mon. Not. R. Astron. Soc*, 000:1–9, 2001.

Reid, B.A. et al. Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. *Monthly Notices of the Royal Astronomical Society*, 404(1):60–85, 5 2010. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2010.16276.X. URL https://academic.oup.com/mnras/article/404/1/60/3101463.

Reid, M.J., Pesce, D.W. and Riess, A.G. An Improved Distance to NGC 4258 and its Implications for the Hubble Constant, 8 2019. ISSN 23318422. URL https://doi.org/10.3847/2041-8213/ab552d.

Renzini, A., Renzini and A. Energetics of stellar populations. *AnPh*, 6:87–102, 1981. ISSN 0003-4169. doi: 10.1051/ANPHYS/198106060087. URL https://ui.adsabs.harvard.edu/abs/1981AnPh....6...87R/abstract.

Rhodes, J.D. et al. The stability of the point-spread function of the advanced camera for surveys on the Hubble Space Telescope and implications for weak gravitational lensing. *Astrophysical journal supplement series, 2007, Vol.172(1), pp.203-218 [Peer Reviewed Journal]*, 172(1):203–218, 9 2007. ISSN 0067-0049. doi: 10.1086/516592. URL http://dx.doi.org/10.1086/516592.

Richards, G.T. et al. Photometric Redshifts of Quasars. *The Astronomical Journal*, 122(3): 1151–1162, 9 2001. ISSN 00046256. doi: 10.1086/322132. URL https://ui.adsabs.harvard.edu/abs/2001AJ....122.1151R/abstract.

Riess, A.G. The Expansion of the Universe is Faster than Expected. *Nature Reviews Physics*, 2(1):10–12, 1 2020. doi: 10.1038/s42254-019-0137-0. URL http://arxiv.org/abs/2001.03624http://dx.doi.org/10.1038/s42254-019-0137-0.

Riess, A.G. et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116(3):1009–1038, 5 1998. doi: 10.1086/300499. URL http://arxiv.org/abs/astro-ph/9805201http://dx.doi.org/10.1086/300499.

Riess, A.G. et al. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics Beyond LambdaCDM. *arXiv*, 3 2019. doi: 10.3847/1538-4357/ab1422. URL http://arxiv.org/abs/1903.07603http://dx.doi.org/10.3847/1538-4357/ab1422.

Risaliti, G. and Lusso, E. Cosmological constraints from the Hubble diagram of quasars at high redshifts. *Nature Astronomy*, 3(3):272–277, 11 2018. URL https://arxiv.org/abs/1811.02590v1.

Robertson, H.P. Kinematics and World-Structure. *The Astrophysical Journal*, 82:284, 11 1935. ISSN 0004-637X. doi: 10.1086/143681. URL https://ui.adsabs.harvard.edu/abs/1935ApJ....82..284R/abstract.

Robertson, H.P. Kinematics and World-Structure III. *The Astrophysical Journal*, 83:257, 5 1936. ISSN 0004-637X. doi: 10.1086/143726. URL https://ui.adsabs.harvard.edu/abs/1936ApJ....83..257R/abstract.

Rocca, J. Bayesian inference problem, MCMC and variational inference | by Joseph Rocca | Towards Data Science, 2019. URL https://towardsdatascience.com/bayesian-inference-problem-mcmc-and-variational-inference-25a8aa9bce29.

Rogozhnikov, A. Hamiltonian Monte Carlo explained, 2016. URL https://arogozhnikov.github.io/2016/12/19/markov_chain_monte_carlo.html.

Ronen, S., Aragón-Salamanca, A. and Lahav, O. Principal component analysis of synthetic galaxy spectra. *Monthly Notices of the Royal Astronomical Society*, 303(2):284–296, 2 1999. ISSN 0035-8711. doi: 10.1046/J.1365-8711.1999.02222.X. URL https://academic.oup.com/mnras/article/303/2/284/1066471.

Ryden, B. Introduction to Cosmology. Technical report, The Ohio State University, 1 2006.

Sadeh, I., Feng, L.L. and Lahav, O. Gravitational Redshift of Galaxies in Clusters from the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey. *Physical Review Letters*, 114(7), 10 2014. doi: 10.1103/PhysRevLett.114.071103. URL http://arxiv.org/abs/1410.5262http://dx.doi.org/10.1103/PhysRevLett.114.071103.

Salvato, M., Ilbert, O. and Hoyle, B. The many flavours of photometric redshifts. *Nature Astronomy*, 3(3):212–222, 5 2018. ISSN 23973366. doi: 10.48550/arxiv.1805.12574. URL https://arxiv.org/abs/1805.12574v2.

Samuroff, S. *Systematic Biases in Weak Lensing Cosmology with the Dark Energy Survey*. PhD thesis, University of Manchester, 2017.

Sandage, A.R. Cosmology: A search for two numbers. *Physics Today*, 23(2):34–41, 2 1970. ISSN 0031-9228. doi: 10.1063/1.3021960. URL /physicstoday/article/23/2/34/427263/Cosmology-A-search-for-two-numbersPrecision.

Sanderson, G. Bayes theorem, the geometry of changing beliefs - YouTube, 2019. URL https://www.youtube.com/watch?v=HZGCoVF3YvM&ab_channel=3Blue1Brown.

Sarkar, S. and Pandey, B. An information theory based search for homogeneity on the largest accessible scale. *Monthly Notices of the Royal Astronomical Society: Letters*, 463(1): L12–L16, 7 2016. doi: 10.1093/mnrasl/slw145. URL http://arxiv.org/abs/1607.06194http://dx.doi.org/10.1093/mnrasl/slw145.

Saulder, C. et al. Calibrating the fundamental plane with SDSS DR8 data. *Astronomy & Astrophysics*, 557:A21, 9 2013. ISSN 0004-6361. doi: 10.1051/0004-6361/201321466. URL https://www.aanda.org/articles/aa/full_html/2013/09/aa21466-13/aa21466-13.htmlhttps://www.aanda.org/articles/aa/abs/2013/09/aa21466-13/aa21466-13.html.

Sawicki, M.J., Lin, H. and Yee, H.K.C. Evolution of the Galaxy Population Based on Photometric Redshifts in the Hubble Deep Field. *The Astronomical Journal*, 113:1, 10 1996. doi: 10.1086/118231. URL http://arxiv.org/abs/astro-ph/9610100http://dx.doi.org/10.1086/118231.

Schechter, P., Schechter and P. An analytic expression for the luminosity function for galaxies. *ApJ*, 203:297–306, 1 1976. ISSN 0004-637X. doi: 10.1086/154079. URL https://ui.adsabs.harvard.edu/abs/1976ApJ...203..297S/abstract.

Scherrer, R.J. Linear velocity fields in non-Gaussian models for large-scale structure. *The Astrophysical Journal*, 390:330, 5 1992. ISSN 0004-637X. doi: 10.1086/171283. URL https://ui.adsabs.harvard.edu/abs/1992ApJ...390..330S/abstract.

Schlafly, E.F. and Finkbeiner, D.P. Measuring Reddening with SDSS Stellar Spectra and Recalibrating SFD. *ApJ*, 737(2):103, 12 2010. doi: 10.1088/0004-637X/737/2/103. URL http://arxiv.org/abs/1012.4804http://dx.doi.org/10.1088/0004-637X/737/2/103.

Schlegel, D.J., Finkbeiner, D.P. and Davis, M. Maps of Dust IR Emission for Use in Estimation of Reddening and CMBR Foregrounds. *ApJ*, 500(2):525–553, 10 1997. doi: 10.1086/305772. URL http://arxiv.org/abs/astro-ph/9710327http://dx.doi.org/10.1086/305772.

Schmitz, M.A. et al. Euclid: Nonparametric point spread function field recovery through interpolation on a graph Laplacian. *Astronomy & Astrophysics*, 636: A78, 4 2020. ISSN 0004-6361. doi: 10.1051/0004-6361/201936094. URL https://www.aanda.org/articles/aa/full_html/2020/04/aa36094-19/aa36094-19.htmlhttps://www.aanda.org/articles/aa/abs/2020/04/aa36094-19/aa36094-19.html.

Schneider, P. et al. A new measure for cosmic shear. *Monthly Notices of the Royal Astronomical Society*, 296(4):873–892, 6 1998. ISSN 0035-8711. doi: 10.1046/J.1365-8711.1998.01422.X. URL https://academic.oup.com/mnras/article/296/4/873/1064847.

Schneider, P. et al. Analysis of two-point statistics of cosmic shear: I. Estimators and covariances. *Astronomy and Astrophysics*, 396(1):1–19, 2002. ISSN 00046361. doi: 10.1051/0004-6361:20021341. URL https://www.researchgate.net/publication/2232569_Analysis_of_two-point_statistics_of_cosmic_shear_I_Estimators_and_covariances.

Schutz, B. *A First Course in General Relativity*. Cambridge University Press, 1985. doi: 10.1017/cbo9780511984181. URL https://www.cambridge.org/core/books/first-course-in-general-relativity/3805425203DD91A7436EF6E5F2082263.

Schutz, B.F. Determining the Hubble constant from gravitational wave observations. *Nature*, 323(6086):310–311, 1986. ISSN 00280836. doi: 10.1038/323310a0. URL https://www.nature.com/articles/323310a0.

Schwarzschild, K., Schwarzschild and Karl. Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. *SPAW*, pages 189–196, 1916. URL https://ui.adsabs.harvard.edu/abs/1916SPAW.....189S/abstract.

Scoccimarro, R. Transients from initial conditions: A perturbative analysis. *Monthly Notices of the Royal Astronomical Society*, 299(4):1097–1118, 10 1998. ISSN 00358711. doi: 10.1046/j.1365-8711.1998.01845.x. URL https://academic.oup.com/mnras/article/299/4/1097/1108910.

Scoville, N. et al. The Cosmic Evolution Survey (COSMOS): Overview. *The Astrophysical Journal Supplement Series*, 172(1):1–8, 9 2007. ISSN 0067-0049. doi: 10.1086/516585. URL https://ui.adsabs.harvard.edu/abs/2007ApJS..172....1S/abstract.

Scranton, R. et al. Detection of Cosmic Magnification with the Sloan Digital Sky Survey. *ApJ*, 633(2):589–602, 11 2005. ISSN 0004-637X. doi: 10.1086/431358. URL https://ui.adsabs.harvard.edu/abs/2005ApJ...633..589S/abstract.

SDSS-III collaboration et al. The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. *Astrophysical Journal, Supplement Series*, 193(2):31, 1 2011. doi: 10.1088/0067-0049/193/2/29. URL http://arxiv.org/abs/1101.1559http://dx.doi.org/10.1088/0067-0049/193/2/29.

Seitz, C. and Schneider, P. Steps towards nonlinear cluster inversion through gravitational distortions III. Including a redshift distribution of the sources. *Astronomy and Astrophysics*, 318(3):687–699, 2 1997.

Sérsic, J.L., Sérsic and L., J. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *BAAA*, 6:41–43, 1963. ISSN 0571-3285. URL https://ui.adsabs.harvard.edu/abs/1963BAAA....6...41S/abstract.

Shajib, A.J. et al. STRIDES: a 3.9 per cent measurement of the Hubble constant from the strong lens system DES J0408-5354. *Monthly Notices of the Royal Astronomical Society*, 494(4):6072–6102, 10 2019. doi: 10.1093/mnras/staa828. URL http://arxiv.org/abs/1910.06306http://dx.doi.org/10.1093/mnras/staa828.

Shariff, H.A. *Application of Bayesian statistics in Supernovae Ia cosmology*. PhD thesis, Imperial College London, 2017. URL https://spiral.imperial.ac.uk/handle/10044/1/63932.

Shectman, S.A. et al. The Las Campanas Redshift Survey. *The Astrophysical Journal*, 470: 172, 10 1996. ISSN 0004-637X. doi: 10.1086/177858. URL https://ui.adsabs.harvard.edu/abs/1996ApJ...470..172S/abstract.

Shirai, S. and Yamazaki, M. Is Gravity the Weakest Force? *Classical and Quantum Gravity*, 38(3), 4 2019. URL http://arxiv.org/abs/1904.10577.

Soares, D.S.L. Newtonian gravitational deflection of light revisited. *ArXiv*, 8 2005. URL https://arxiv.org/abs/physics/0508030v4.

Sofue, Y. and Rubin, V. Rotation Curves of Spiral Galaxies. *Annual Review of Astronomy and Astrophysics*, 39(1):137–174, 10 2000. doi: 10.1146/annurev.astro.39.1.137. URL http://arxiv.org/abs/astro-ph/0010594http://dx.doi.org/10.1146/annurev.astro.39.1.137.

Solà, J. Cosmological constant and vacuum energy: Old and new ideas. *Journal of Physics: Conference Series*, 453(1), 2013. ISSN 17426596. doi: 10.1088/1742-6596/453/1/012015. URL https://www.researchgate.net/publication/237054399_Cosmological_constant_and_vacuum_energy_Old_and_new_ideas.

Soldner, J.G.v. *Ueber die Ablenkung eines Lichtstrals von seiner geradlinigen Bewegung, durch die Attraktion eines Weltk¨orpers, an welchem er nahe vorbei geht*. bey dem Verfasser, Berlin, 1804.

Sousbie, T. et al. The Three-dimensional Skeleton of the SDSS. *The Astrophysical Journal*, 672(1):L1–L4, 1 2008. ISSN 0004-637X. doi: 10.1086/523669. URL https://iopscience.iop.org/article/10.1086/523669https://iopscience.iop.org/article/10.1086/523669/meta.

Sparnaay, M.J. Attractive forces between flat plates [6], 1957. ISSN 00280836. URL https://www.nature.com/articles/180334b0.

Spitzer, L. and L. Physics of Fully Ionized Gases. *pfig*, 1956. URL https://ui.adsabs.harvard.edu/abs/1956pfig.book.....S/abstract.

Springob, C.M. et al. The 6dF Galaxy Survey: peculiar velocity field and cosmography. *Monthly Notices of the Royal Astronomical Society*, 445(3):2677–2697, 12 2014. ISSN 1365-2966. doi: 10.1093/mnras/stu1743. URL http://academic.oup.com/mnras/article/445/3/2677/1036663/The-6dF-Galaxy-Survey-peculiar-velocity-field-and.

SubbaRao, M.U. et al. Luminosity Functions from Photometric Redshifts I: Techniques. *AJ*, 112:929, 6 1996. doi: 10.1086/118066. URL http://arxiv.org/abs/astro-ph/9606075http://dx.doi.org/10.1086/118066.

Sutter, P.M. et al. VIDE: The Void IDentification and Examination toolkit. *Astronomy and Computing*, 9:1–9, 6 2014a. URL http://arxiv.org/abs/1406.1191.

Sutter, P.M. et al. Voids in the SDSS DR9: observations, simulations, and the impact of the survey mask. *Monthly Notices of the Royal Astronomical Society*, 442(4):3127–3137, 8 2014b. ISSN 0035-8711. doi: 10.1093/MNRAS/STU1094. URL https://academic.oup.com/mnras/article/442/4/3127/1354064.

Takada, M. and White, M. Tomography of lensing cross power spectra. *The Astrophysical Journal*, 601(1):L1–L4, 11 2003. doi: 10.1086/381870. URL http://arxiv.org/abs/astro-ph/0311104http://dx.doi.org/10.1086/381870.

Tanaka, M. et al. Photometric redshifts for Hyper Suprime-Cam Subaru Strategic Program Data Release 1. *Publications of the Astronomical Society of Japan*, 70(SP1):9–10, 1 2018. ISSN 0004-6264. doi: 10.1093/PASJ/PSX077. URL https://academic.oup.com/pasj/article/70/SP1/S9/4494086.

The IceCube Collaboration et al. The IceCube Neutrino Observatory - Contributions to ICRC 2015 Part IV: Searches for Dark Matter and Exotic Particles. *J. Sandroos*, 30, 10 2015. URL http://arxiv.org/abs/1510.05226.

The MACHO collaboration et al. The MACHO Project: Microlensing Results from 5.7 Years of LMC Observations. *The Astrophysical Journal*, 542(1):281–307, 1 2000. doi: 10.1086/309512. URL http://arxiv.org/abs/astro-ph/0001272http://dx.doi.org/10.1086/309512.

Tollerud, E.J. et al. The SPLASH Survey: Spectroscopy of 15 M31 Dwarf Spheroidal Satellite Galaxies. *Astrophysical Journal*, 752(1), 12 2011. doi: 10.1088/0004-637X/752/1/45. URL http://arxiv.org/abs/1112.1067http://dx.doi.org/10.1088/0004-637X/752/1/45.

Treschman, K.J. Early Astronomical Tests of General Relativity: the anomalous advance in the perihelion of Mercury and gravitational redshif. *Asian Journal of Physics*, 23, 2014.

Troxel, M.A. and Ishak, M. The Intrinsic Alignment of Galaxies and its Impact on Weak Gravitational Lensing in an Era of Precision Cosmology. *Physics Reports*, 558:1–59, 7 2014. doi: 10.1016/j.physrep.2014.11.001. URL http://arxiv.org/abs/1407.6990http://dx.doi.org/10.1016/j.physrep.2014.11.001.

Tucker, W. et al. E0657-56: A CONTENDER FOR THE HOTTEST KNOWN CLUSTER OF GALAXIES. Technical report, 1998.

Tully, R. and Fisher, J. A New method of determining distances to galaxies. *Astron.Astrophys.*, 54:661–673, 1977.

Tversky, A. and Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, 9 1974. ISSN 00368075. doi: 10.1126/SCIENCE.185.4157.1124. URL https://www.science.org/doi/10.1126/science.185.4157.1124.

Vaccari, M. et al. in Science from the Next Generation Imaging and Spectroscopic Surveys. page 49, 2012.

Vallinotto, A. Using CMB lensing to constrain the multiplicative bias of cosmic shear. *Astrophysical Journal*, 759(1), 10 2011. doi: 10.1088/0004-637X/759/1/32. URL http://arxiv.org/abs/1110.5339http://dx.doi.org/10.1088/0004-637X/759/1/32.

Van Waerbeke, L. Shear and magnification: cosmic complementarity. *Monthly Notices of the Royal Astronomical Society*, 401(3):2093–2100, 1 2010. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2009.15809.X. URL https://academic.oup.com/mnras/article/401/3/2093/1099616.

Van Waerbeke, L. and Mellier, Y. Gravitational Lensing by Large Scale Structures: A Review. *ArXiv*, 5 2003. URL https://arxiv.org/abs/astro-ph/0305089v2.

Van Waerbeke, L. et al. Likelihood analysis of cosmic shear on simulated and VIRMOS-DESCART data. *A&A*, 393:369–379, 2002. doi: 10.1051/0004-6361:20020932. URL http://dx.doi.org/10.1051/0004-6361:20020932.

Vanzella, E. et al. Photometric redshifts with the Multilayer Perceptron Neural Network: Application to the HDF-S and SDSS. *Astronomy & Astrophysics*, 423(2):761–776, 8 2004. ISSN 0004-6361. doi: 10.1051/0004-6361:20040176. URL https://www.aanda.org/articles/aa/abs/2004/32/aa3658/aa3658.html.

Venemans, B.P. et al. First discoveries of z \sim 6 quasars with the Kilo-Degree Survey and VISTA Kilo-Degree Infrared Galaxy survey. *Monthly Notices of the Royal Astronomical Society*, 453(3):2259–2266, 11 2015. ISSN 0035-8711. doi: 10.1093/MNRAS/STV1774. URL https://academic.oup.com/mnras/article/453/3/2259/1079051.

Voigt, L.M. and Bridle, S.L. Limitations of model-fitting methods for lensing shear estimation. *Monthly Notices of the Royal Astronomical Society*, 404(1):458–467, 5 2010. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2010.16300.X. URL https://academic.oup.com/mnras/article/404/1/458/3101597.

Wagh, S.M. Some fundamental issues in General Relativity and their resolution. *ArXiv*, 2 2004. URL http://arxiv.org/abs/gr-qc/0402003.

Walcher, C.J. et al. Fitting the integrated Spectral Energy Distributions of Galaxies. *Astrophysics and Space Science*, 331(1):1–51, 8 2010. doi: 10.1007/s10509-010-0458-z. URL http://arxiv.org/abs/1008.0395http://dx.doi.org/10.1007/s10509-010-0458-z.

Ward, B. Tilt Classifications in Perfect Fluid Cosmology. *Physics Capstone Projects*, 5 2017. URL https://digitalcommons.usu.edu/phys_capstoneproject/54.

Weinberg, D.H. et al. Cold dark matter: controversies on small scales. *Proceedings of the National Academy of Sciences of the United States of America*, 112(40):12249–12255, 6 2013. doi: 10.1073/pnas.1308716112. URL http://arxiv.org/abs/1306.0913http://dx.doi.org/10.1073/pnas.1308716112.

Weinberg, S. The cosmological constant problem. *Reviews of Modern Physics*, 61(1):1, 1 1989. ISSN 00346861. doi: 10.1103/RevModPhys.61.1. URL https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.61.1.

Weinberg, S. and Wagoner, R.V. Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity . *Physics Today*, 26(6):57–58, 6 1973. ISSN 0031-9228. doi: 10.1063/1.3128097. URL http://physicstoday.scitation.org/doi/10.1063/1.3128097.

Weinstein, M.A. et al. An Empirical Algorithm for Broadband Photometric Redshifts of Quasars from the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series*, 155(2):243–256, 12 2004. ISSN 0067-0049. doi: 10.1086/425355/FULLTEXT/. URL https://iopscience.iop.org/article/10.1086/425355https://iopscience.iop.org/article/10.1086/425355/meta.

Weymann, R.J., Carswell, R.F. and Smith, M.G. Absorption Lines in the Spectra of Quasistellar Objects. *Annual Review of Astronomy and Astrophysics*, 19(1):41–76, 9 1981. ISSN 0066-4146. doi: 10.1146/annurev.aa.19.090181.000353. URL https://ui.adsabs.harvard.edu/abs/1981ARA&A..19...41W/abstract.

Williams, R.E. and team, t.H. The Hubble Deep Field: Observations, Data Reduction, and Galaxy Photometry. *The Astronomical Journal*, 112:1335, 7 1996. doi: 10.1086/118105. URL http://arxiv.org/abs/astro-ph/9607174http://dx.doi.org/10.1086/118105.

Willman, B. et al. A New Milky Way Dwarf Galaxy in Ursa Major. *The Astrophysical Journal*, 626(2):L85–L88, 3 2005. doi: 10.1086/431760. URL http://arxiv.org/abs/astro-ph/0503552http://dx.doi.org/10.1086/431760.

Wittman, D. WHAT LIES BENEATH: USING p(z) TO REDUCE SYSTEMATIC PHOTOMETRIC REDSHIFT ERRORS. *The Astrophysical Journal*, 700 (2):L174, 7 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/700/2/L174. URL https://iopscience.iop.org/article/10.1088/0004-637X/700/2/L174https://iopscience.iop.org/article/10.1088/0004-637X/700/2/L174/meta.

Wojtak, R., Hansen, S.H. and Hjorth, J. Gravitational redshift of galaxies in clusters as predicted by general relativity. *Nature*, 477(7366):567–569, 9 2011. doi: 10.1038/nature10445. URL http://arxiv.org/abs/1109.6571http://dx.doi.org/10.1038/nature10445.

Wolf, C. Bayesian photometric redshifts with empirical training sets. *Monthly Notices of the Royal Astronomical Society*, 397(1):520–533, 7 2009. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2009.14953.X. URL https://academic.oup.com/mnras/article/397/1/520/1007663.

Wong, K.C. et al. H0LiCOW XIII. A 2.4% measurement of H_{0} from lensed quasars: 5.3σ tension between early and late-Universe probes. *Monthly Notices of the Royal Astronomical Society*, 498(1):1420–1439, 7 2019. doi: 10.1093/mnras/stz3094. URL http://arxiv.org/abs/1907.04869http://dx.doi.org/10.1093/mnras/stz3094.

Wright, A.H. et al. KiDS+VIKING-450: A new combined optical and near-infrared dataset for cosmology and astrophysics. *Astronomy & Astrophysics*, 632:A34, 12 2019. ISSN 0004-6361. doi: 10.1051/0004-6361/201834879. URL https://www.aanda.org/articles/aa/full_html/2019/12/aa34879-18/aa34879-18.htmlhttps://www.aanda.org/articles/aa/abs/2019/12/aa34879-18/aa34879-18.html.

Wright, A.H. et al. Photometric redshift calibration with self-organising maps. *Astronomy & Astrophysics*, 637:A100, 5 2020. ISSN 0004-6361. doi: 10.1051/0004-6361/201936782. URL https://www.aanda.org/articles/aa/full_html/2020/05/aa36782-19/aa36782-19.htmlhttps://www.aanda.org/articles/aa/abs/2020/05/aa36782-19/aa36782-19.html.

Yèche, C. et al. Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100. *Journal of Cosmology and Astroparticle Physics*, 2017(6):047, 6 2017. ISSN 14757516. doi: 10.1088/1475-7516/2017/06/047. URL https://iopscience.iop.org/article/10.1088/1475-7516/2017/06/047https://iopscience.iop.org/article/10.1088/1475-7516/2017/06/047/meta.

Young, B.L. A survey of dark matter and related topics in cosmology. *Front. Phys*, 12(2): 121201, 2017. doi: 10.1007/s11467-016-0583-4. URL www.springer.com/11467.

Zaidouni, F. Gravitational Lensing. Technical report, University of Rochester, 12 2018.

Zeger, S.L. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy. *Physics Today*, 65(7):54, 7 2012. ISSN 0031-9228. doi: 10.1063/PT.3.1646. URL https://physicstoday.scitation.org/doi/abs/10.1063/PT.3.1646.

Zel'Dovich, Y.B., Zel'Dovich and B., Y. Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations. *A&A*, 500:13–18, 1970. ISSN 0004-6361. URL https://ui.adsabs.harvard.edu/abs/1970A&A.....5...84Z/abstract.

Zhang, P. Self calibration of gravitational shear-galaxy intrinsic ellipticity correlation in weak lensing surveys. *Astrophysical Journal*, 720(2):1090–1101, 11 2008. doi: 10.1088/0004-637X/720/2/1090. URL http://arxiv.org/abs/0811.0613http://dx.doi.org/10.1088/0004-637X/720/2/1090.

Zhang, T. et al. Impact of point spread function higher moments error on weak gravitational lensing. *Mon.Not.Roy.Astron.Soc.*, 510(2):1978–1993, 2 2022. ISSN 13652966. doi: 10.1093/MNRAS/STAB3584.

Zwicky, F. On the Masses of Nebulae and of Clusters of Nebulae. *The Astrophysical Journal*, 86:217, 10 1937. ISSN 0004-637X. doi: 10.1086/143864. URL https://ui.adsabs.harvard.edu/abs/1937ApJ....86..217Z/abstract.