Title
Photometric Redshift Systematics In Optical Cosmology Surveys

Author
Hasan, Imran Syed

Publication Date
2021

Peer reviewed|Thesis/dissertation

Photometric Redshift Systematics In Optical Cosmology Surveys

By

IMRAN SYED HASAN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

J. Anthony Tyson (Chair)

David Wittman

Marusa Bradac

Committee in Charge

2021

Contents

Photometric Redshift Systematics In Optical Cosmology Surveys

Abstract

Weak lensing cosmology surveys measure the light from distant galaxies as it undergoes a series of distortions, suffering from deflections by both the large scale structure of the universe and individual galaxies. As the number of galaxies used to measure cosmology have increased into the tens of millions, the myriad of pernicious systematic errors to dominate over random errors. Over time and through different generations of surveys, understanding of leading sources of systematic error—like model inaccuracies of the Point Spread Function (PSF), observation strategies, the brighter-fatter effect, and differed charge transfer—has matured, leaving other sources of systematic error to become dominant. One source of systematic error that still demands the community's attention is accurately modeling the redshifts of weak lensing galaxies. Perhaps the most secure way to measure galaxy redshifts is by measuring spectroscopic redshifts. Obtaining spectra for millions of faint galaxies that make up weak lensing surveys is intractable, however. Instead, weak lensing surveys traditionally estimate galaxy redshifts by using their flux as measured through broad band photometric filters, resulting in photometric redshifts. In this work we will examine the effects of photometric redshift uncertainty have our ability to constrain cosmological parameters. First, we will utilize the Deep Lensing Survey to study the effect that the error in estimated redshift distribution widths, $n(z)$, has on cosmology inference using the joint combination of galaxy clustering, galaxy-galaxy lensing, and cosmic shear. We present two end—to—end analyses from the catalog level to parameter estimation. We produce an initial cosmological inference using fiducial tomographic redshift bins derived from photometric redshifts, then compare this with a result where the redshift bins are empirically corrected using a set of spectroscopic redshifts. We find that the derived parameter $S_8 \equiv \sigma_8(\Omega_m/.3)^{1/2}$ decreases from $0.841^{+0.062}_{-0.061}$ to $0.810^{+0.039}_{-0.031}$ upon correcting the n(z) errors in the second method. Second, we study various approaches to estimate photometric redshift of the Dark Energy Science Collaboration (DESC)'s Data Challenge (DC) 2 catalogs and see how the blended objects affect the result of photo-z estimations. We also explore the subsequent effect these errors have on our ability to measure the dark matter equation of state, w, and find if used alone, blending related photo-z errors can induce up to a 160% error in w.

Acknowledgments and Thanks

To my parents, for their constant support. To my cohort, for sharing the journey. To my office mates and group members, for the late night conversations on science and other shenanigans. To my committee members, their feedback greatly improved this work. To Tony, for his guidance and wisdom

Chapter 1

Introduction

Shortly after the Big Bang, the universe underwent an era of inflation, expanding rapidly and imparting small inhomogeneities in the universe. These primordial fluctuations planted the seeds for fluctuations in the matter density distribution in universe that we observe today, forming large dark matter halos that host galaxies. Several important phenomena connect the very early universe to the one observed today. In particular, the accelerating expansion of the universe has strongly influenced the formation history of dark matter halos and their galaxies that are ultimately produced. Gravity acts to drive mass towards initially over dense regions, causing it to accumulate and become more over dense still, creating what we will refer to informally as 'structures.' However, the growth of structure can be impeded or even suppressed entirely by the late time accelerating expansion of the universe. Two possible explanations for late time acceleration are alternate theories of gravity, or a form of vacuum energy referred to broadly as *dark energy*.

Dark energy may be the result of a negative pressure to the universe. It can be parameterized by its equation of state parameter (or 'equation of state' for shorthand) $w_{DE} = p_{DE}/\rho_{DE}$, the ratio between its pressure and density. Dark energy can be modeled as a 'cosmological constant', Λ, in the case that $w_{DE} = -1$, that is tacked onto the stress energy tensor in Einstein's field equations for general relativity. To see this, one can start from the continuity equation and the Euler equation, and show that

$$\frac{\partial \rho}{\partial t} + \frac{\dot{a}}{a}[3\rho + 3P] = 0 \tag{1.0.1}$$

where ρ represents the different species of energy density and P represents pressure. Taking $\rho_{DE} =$

$-P_{DE}$, or equivalently $w_{DE} = -1$, we are left with $\frac{\partial \rho}{\partial t} = 0$, implying the energy density for dark energy is constant and does not change with time. Λ contributes a constant energy density to every point in space time and is the most abundant form of matter-energy in the today's universe.

Even if a cosmological constant scenario has taken its place in the center stage of the most popular cosmological theories, investigations into alternative theories of dark energy, and experimental efforts to rule these out are still underway. For example, if dark energy is due to a cosmological constant this would imply $w_{DE} = -1$. Therefore, observational measures of w_{DE} can distinguish between a cosmological constant form of dark energy and other scenarios. One avenue of pursuing this is to measure any time variation in $w_{DE}(t)$, as departures away from $w = -1$ provide direct constraints. This can be seen by referring to Eq 1.0.1. In the case where w_{DE} is not -1, we are left with a time dependent equation for ρ_{DE}

The time evolution of matter structures are phenomenologically tied to the expansion of the universe, and dark energy by extension. Consequently, measuring matter structure formation at different cosmic epochs ultimately indirectly probes $w(t)_{DE}$. This can provide an opportunity to falsify, or at the least stress test, the case that dark energy is due to a cosmological constant.

Since the universe's expansion accelerates at late times, the recent universe provides a powerful theater in which to investigate dark energy. Because analysis of matter structures anchor our examinations of dark energy, we turn to astronomical imaging surveys of the recent universe. Provided our imaging survey can probe different epochs, we can measure structure at different snapshots in the universe's history. Comparing structure variations from snapshot to snapshot then provides a measure of its growth, and dark energy by extension.

Typically, observed galaxies are used as tracers of the matter field. These luminous galaxies tell part of the story of matter structures, as they live in a dark matter dominated scaffolding. Because luminous galaxies are intrinsically tied to the underlying dark matter in the universe, studying their clustering over time informs us of the evolution of the matter field as it is influenced by gravity and the expansion of the universe.

In addition to using the positions and distances of galaxies to trace out the large scale structure they lie in, the shapes of galaxies can also be used to trace out large matter structures. Gravitational lensing, the phenomena in which the shapes of background galaxies are distorted into coherent tangential alignment by the gravitational field of a foreground galaxy, traces out the mass profiles of matter halos. Using this effect to characterize the mass profiles of galaxies and the dark matter they are embedded in can allow us to examine them as a function of time. Since this formation is influenced by the expansion of the universe and gravity, it gives us leverage in understanding dark energy and dark matter.

With an eye towards experimentally measuring the clustering of galaxies and gravitational lensing by galaxies, and using our measurements to parameterize the ΛCDM model, we will present fundamental principles upon which such measurements are built upon. Following will be a brief review of cosmology fundamentals, the growth of cosmic structure, statistical measures of structure, gravitational lensing, and detecting and measuring galaxy shapes.

1.1 Cosmology Fundamentals

We set the stage by providing a background into General Relativity, dark matter, dark energy, and the central roles they occupy in the most favored cosmological theory, ΛCDM.

Einstein's theory of general relativity has a firm foundation on top of which the ΛCDM cosmological model is constructed. In this framework, the stress energy tensor, $T_{\mu\nu}$, and cosmological constant, Λ, are related to the curvature of space time in the following form

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu} - \Lambda g_{\mu\nu} \tag{1.1.1}$$

Where G is Newton's gravitational constant, and $R_{\mu\nu}$ (R) and $g_{\mu\nu}$ are the Ricci Tensor (scalar), and metric respectively, describing the geometry of the universe. The stress-energy tensor describes the flux and density of energy and momentum contained in the universe. In the ΛCDM picture, $T_{\mu\nu}$ includes contributions not only from baryonic matter–the familiar form of matter comprised of protons, neutrons, electrons and the like–but also dark matter. Dark matter interacts with gravity, and does not measurably interact with light. The cosmological constant, Λ, can be thought of as an energy density which provides a negative pressure in the universe, driving its expansion at late times.

If we assume matter-energy at large scales is homogeneous and isotropic, we arrive at a solution to Einstein's field equations referred to as the FRWL metric and first FRWL equation

$$ds^2 = -dt^2 + a^2(t)\Big(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\Big) \tag{1.1.2}$$

$$H^2 + \frac{k}{a^2} = \frac{8\pi G_N}{3}\rho_{tot} \tag{1.1.3}$$

In above in equation 1.1.2, ds is the invariant, k parameterizes the curvature of the universe, a is the scale factor-defined to have a value of 0 at the big bang and a value of 1 today. The second equation describes the time evolution of the scale factor, where $H = \dot{a}/a$ is the Hubble parameter, and ρ_{tot} is the total matter-energy density, which we may subdivide into different species,

$$\Omega_i = \frac{\rho_i}{\rho_c} \tag{1.1.4}$$

where

$$\Omega = \sum_i \Omega_i, \tag{1.1.5}$$

$$\rho_c = \frac{3H^2}{8\pi G} \tag{1.1.6}$$

The different species are Ω_M for non relativistic matter-energy, Ω_R for relativistic matter-energy, Ω_k for energy associated with the curvature of the universe, and Ω_Λ for the dark energy content. In the ΛCDM paradigm, k is typically taken to be zero, which corresponds to a flat universe. Today, Ω_R is at a scale of 10^{-4}, and thus the mass-energy budget is dominated by Ω_M and Ω_Λ, matter and dark energy densities respectively.

Ω_M can be further subdivided into dark matter and baryonic matter components. Measurements of anisotropies in the cosmic microwave background , measurements of quasar spectra, and measurements of luminous mass in galaxy clusters puts baryonic matter, Ω_b, to be $\sim$ a factor of 5 less abundant than dark matter. Returning our attention to the equation for the Hubble parameter, it is apparent that the expansion history of the universe depends on the matter-energy budget. As a result, during late times when Ω_Λ makes up $\sim$ 70 percent of the energy budget, any probe of phenomena that depend on H is also a probe of dark energy.

1.2 Growth of Structure

The variations in the primordial matter density field that were imprinted on the universe by the inflationary era ultimately evolve to create over dense regions in the universe that host galaxies and galaxy clusters observed in the recent universe. These galaxy scale and larger 'peaks' in the matter density field are what we refer to when discussing cosmic structure. Because we are concerned with the evolution of matter density perturbations, it is useful to consider the density contrast

$$\delta = (\rho(x,t) - \bar{\rho}(x,t))/\bar{\rho}(x,t) \tag{1.2.1}$$

which characterizes local over and under densities in the matter density field. In the following discussion, we will work in the linear regime, where $\delta \ll 1$ (Elvin-Poole et al., 2018). The time and spatial dependence in $\delta(x,t)$ can be separated out by writing

$$\delta(x,t) = \delta_0 D(t) \tag{1.2.2}$$

Where δ_0 is the initial density contrast field, and $D(t)$ is the growth function, which evolves δ_0 in time by linearly modulating its spatial shape (Dodelson, S., 2003). The growth function captures the time dependent effects the Hubble expansion on the evolution of δ_0. The effect of the growth function is a strong function of how the energy budget of the universe is divided up. For example, in the radiation dominated era, perturbations with comoving scales smaller than the horizon do not grow. During a matter dominated era where $\Omega_m \approx 1$, $D_+ \propto a$. Finally, the growth of perturbations is suppressed in a dark energy dominated era (Kilbinger, 2015).

We develop this further by writing the equation of motion for $\delta(x,t)$. Recalling that the matter field is embedded in an expanding universe, applying Poisson's equation in matter, and enforcing continuity of matter and momentum leads to

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\rho\delta = 0 \tag{1.2.3}$$

In the last two terms we can see the competing effects of cosmic expansion and gravity (Elvin-Poole et al., 2018). The middle term is a dampening term-the accelerating expansion smoothing out density perturbations. The last term describes gravity's role in driving matter to initially highly over dense regions, creating structures. All together, the equation of motion shows there is an intimate relationship between the initial density perturbations and their evolution-which is a strong function the expansion history of the universe. As a consequence, if we can characterize the matter field at different times in the universe's history, we will be able to back out the expansion history of the universe.

1.3 Two Point Correlations

We rely on statistical measures to ultimately constrain the growth of structures in the universe observationally. We begin this section by first developing these statistics formally, and close by emphasizing their utility highlighting what they tell us about the cosmological evolution of structure.

We will make repeated use of the 2-point correlation function, ξ, which measures the correlation between two fields. The 2-point correlation function for two fields which are assumed to be homogeneous and isotropic, ϕ and ψ is given by

$$\xi(\vec{r})_{\phi\psi} = \int \phi(\vec{r}\,')\psi(\vec{r}\,'-\vec{r}\,)\mathrm{d}^3 r' \tag{1.3.1}$$

As an example, if we observe a peak in δ, we may ask what the likelihood is of finding another peak with separation $\vec{r}$ away. If P_V is the likelihood of finding a peak in a volume V, the likelihood of finding a neighboring peak $\vec{r}$ away is given by $P_V^2[1+\xi_{\delta\delta}]$ (Dodelson, S., 2003). If $\xi_{\delta\delta}$ is zero, than the peaks do not 'know' about each other, and the positions of peaks are statistically independent of each other. Colloquially, the auto-correlation of δ, $\xi_{\delta\delta}$ tells us to what extent the shape of the peaks and troughs in the field resembles itself as a function of separation between peaks and troughs.

We may also define the mass variance as the 2-point auto-correlation evaluated at zero, $\sigma^2 = \xi_{\delta\delta}(0)$. σ is then equivalent to the root mean square of mass fluctuations and gives a sense of the variability of over/under densities. More colloquially it characterizes how 'peaky' δ is. Traditionally, authors have considered the parameter σ_8, where the variance is evaluated after the primordial contrast field is convolved with a spherical top hat function of radius 8 megaparsecs, and linearly evolved to the present day. It is also useful to consider properties of the density contrast in Fourier space,

$$\langle \delta_k \delta^*_{k'} \rangle = 2\pi\delta(k-k')P(k) \tag{1.3.2}$$

where we have defined $P(k)$ as the *power spectrum*, form which we can now derive aforementioned statistics (Dodelson, S., 2003). The two-point correlation function is given by the relation

$$\xi = \frac{1}{2\pi^2}\int \mathrm{d}k k^2 P(k)\frac{\sin(kr)}{kr} \tag{1.3.3}$$

where we have used Perceval's theorem. σ_8 is given by the relation

$$\sigma_8 = \frac{1}{2\pi^2}\int \mathrm{d}k k^2 P(k)\widetilde{W}(k)$$

where we have used the convolution theorem, and $\widetilde{W}$ is the Fourier transform of the spherical top hat (Schneider, P., 2006).

$\xi_{\delta\delta}$ and σ_8 statistically describe the fluctuations in matter, and quantify structures in the universe.

However, the bulk of matter is in the form of dark matter. We can make statistical measures of quantities that are correlated with the underlying matter density field to elucidate our understanding of the dark matter field. Luminous galaxies, conveniently, are well suited to this task. Their relative high concentration of baryonic content makes them detectable, and their positions trace out the densest regions of the matter density field.

There are two two-point correlation functions of interest we will use in this work to probe the matter field. The first is the angular two point correlation function, $w(\theta)$ which correlates the angular position of galaxies on the sky, separated by an angle θ. Because this uses the most over dense regions of the matter field to probe, it will give us a biased estimate of the matter field, $\xi_{gg} = b^2\xi_{\delta\delta}$ where we have defined the linear galaxy bias, b. The second probe is galaxy-galaxy lensing, where we correlate the positions of foreground galaxies with the background convergence field. Because we again are correlating the positions of the most over dense regions in the matter field, this probe will suffer from the galaxy bias, $\xi_{g\kappa} = b\xi_{\delta\delta}$, this time linearly (Mandelbaum, 2018). More discussion is in order on the lensing phenomena, and we discuss it in detail in the following section.

1.4 Gravitational Lensing Fundamentals

Dark matter is approximately 5 times more abundant in the universe than baryonic matter, and any measure of matter structures on cosmological scales must take its contribution to the matter budget of the universe into account. While dark matter does not appear to interact with light, it does interact with gravity. By characterising the lensing effect that dark matter has on light, one can constrain the nature of dark matter by extension.

The canonical example to understanding gravitational lensing is to begin with a source at an angular diameter distance D_s from the observer, and a lens—whose mass will curve its surrounding space time, deflecting the source's light towards the observer's eye—fixed in the foreground 'lens plane' at an angular diameter distance distance D_l from the observer. Let the angular diameter distance from the lens plane to the source plane be D_{ls}.

If the lens and source are point sources, the deflection angle, α, is given by

$$\alpha = \frac{4GM}{c^2\xi} \tag{1.4.1}$$

where M is the mass of the lens and ξ here is the impact parameter, not to be confused with the two point correlation function. If we generalize the lens to be an extended body with density $\rho(\xi, r)$ where r is the axis along which light from the source propagates, the net deflection angle is given by the sum of deflection angles each infinitesimal bit of mass from the lens imparts on the source's photons (Schneider, P. et al, 2006).

$$\alpha = \frac{4G}{c^2}\int \rho(\xi, r)\frac{\xi - \xi'}{|\xi - \xi'|^2}\mathrm{d}\xi^2\mathrm{d}r \tag{1.4.2}$$

Since the spatial extent of the lens is small compared to the distances being considered, we will work in the thin lens limit, where we consider the 2 dimensional projected surface density of the lens. We may first define the surface density as

$$\Sigma(\xi) = \int \rho(\xi, r)\mathrm{d}r \tag{1.4.3}$$

and subsequently define the dimensionless surface mass density, or convergence

$$\kappa = \frac{\Sigma(D_l\theta)}{\Sigma_{crit}} \tag{1.4.4}$$

where the angular position $\xi = D_l\theta$

$$\Sigma_{crit} = \frac{c^2 D_s}{4\pi G D_l D_{ls}} \tag{1.4.5}$$

We may then define the lensing potential

$$\psi(\theta) = \int \mathrm{d}\theta'^2 \kappa(\theta^{'})\ln|\theta - \theta^{'}| \tag{1.4.6}$$

and redefine the deflection angle as $\alpha = \nabla\psi$ (Kilbinger, 2015) Then using the lens equation, $\vec{\beta} = \vec{\theta} - \frac{D_{ds}}{D_s}\vec{\alpha}$. We work in the limiting case where sources are small enough that their κ are constant over the face of the source, we can define $\theta_{ul} = A\theta_l$ and see the mapping from lensed to unlensed coordinates is a linear transformation, where we have defined the Jacobi matrix $A_{ij} = \delta_{ij} - \partial_i\partial_j\psi$ (Mandelbaum, 2018)

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} \tag{1.4.7}$$

where $\gamma_1 = \frac{1}{2}(\psi_{,11} - \psi_{,22})$ and $\gamma_2 = \psi_{,12}$ and $2\kappa = \nabla^2\psi$. Here the subscripts indicate partial derivatives with respect to the ij components of θ. The 2 component complex shear, $\gamma_1 + i\gamma_2 = \gamma$ describes the stretch of the source galaxy in the weak lensing limit while the convergence, κ describes the isotropic change of size and brightness of the source galaxy.

We can factor out (1 - κ) from the matrix, because this affects the isotropic size of the source, and not the shape. Weak lensing cosmological probes are based on *observed* galaxy shapes, and therefore we are measuring the *reduced shear* (Kilbinger, 2015)

$$g = \frac{\gamma}{1-\kappa} \tag{1.4.8}$$

and not the shear γ. In the weak lensing limit κ and γ are much smaller than unity, and $g \approx \gamma$.

1.5 Estimating Shear From Galaxies

We can ascribe a complex intrinsic ellipticity, ϵ^s to each galaxy. Under weak lensing, the ellipticity is transformed. If the *observed* galaxy has an ellipticity with semi-major and semi-minor axis a and b respectively, we may write the observed ellipticity as $\epsilon = (a - b)\exp(\pi i\phi)/(a + b)$ with position angle ϕ. The observed and intrinsic ellipticity are related to one another by the reduced complex shear by

$$\epsilon = \frac{\epsilon^s + g}{1 + \epsilon^s g^*} \tag{1.5.1}$$

provided that $|g| < 1$. In the weak lensing limit, this reduces to $\epsilon = \epsilon^s + \gamma$ (Kilbinger, 2015). Because galaxies do not have any preferred orientation ϕ, $\langle\epsilon^s\rangle = 0$ and

$$\langle\epsilon\rangle \approx g \tag{1.5.2}$$

Thus by averaging over an ensemble of galaxy shapes we can obtain an estimate of the reduced shear.

We should note, however, that we are not interested in the shapes of individual galaxies, we are interested in the ensemble weighted average of shapes, since that is an estimator of the reduced shear. In fact, the "shape" of a galaxy is ill-defined. Often galaxies are not elliptical in shape, have complex morphology or elliptical gradients. Additionally, the measured shape is a function of the radial weight function used to calculate the ellipticities, which varies from algorithm to algorithm(Mandelbaum, 2018).

1.6 Galaxy-Galaxy Lensing

Galaxy-galaxy lensing is quantified as the tangential shear $\langle\gamma_t\rangle$ of background galaxies around a lens galaxy. It is the projected mass within an aperture on the sky. The mean convergence inside a circular aperture with angular radius θ is

$$\bar{\kappa}(\leq\theta) = \frac{1}{\pi\theta^2}\int_{\theta'<\theta} \mathrm{d}\theta' \kappa(\theta') = \frac{2}{\theta^2}\int_0^\theta \mathrm{d}\theta' \langle\kappa(\theta')\rangle \tag{1.6.1}$$

where we have averaged out in the azimuthal direction, defining $\langle\kappa(\theta')\rangle = \frac{1}{2}\pi \int_0^{2\pi} \kappa(\theta', \phi)\mathrm{d}\phi$ (Kilbinger, 2015)

Now, since $2\kappa = \nabla^2\psi$, we can use the divergence theorem and write the projected surface density

$$\bar{\kappa}(\leq \theta) = \frac{1}{2\pi\theta} \int_0^{2\pi} \mathrm{d}\phi \partial_\theta \psi(\phi, \theta) \tag{1.6.2}$$

Now, using another derivative in the angular radial direction will introduce two derivatives on ψ and we may use the relation $2\kappa = \nabla^2\psi$ once more.

$$\frac{\partial[\theta\bar{\kappa}(\leq \theta)]}{\partial\theta} = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\phi \partial_\theta \partial_\theta \psi(\phi, \theta) = \langle\kappa\rangle(\theta) - \langle\gamma_t\rangle(\theta) \tag{1.6.3}$$

on the other hand, taking the derivative on 1.6.1 gives us

$$\frac{\partial[\theta\bar{\kappa}(\leq \theta)]}{\partial\theta} = -\bar{\kappa}(\leq \theta) + 2\langle\kappa\rangle(\theta) \tag{1.6.4}$$

equating these two finally gives us

$$\langle\gamma_t\rangle = \bar{\kappa}(\leq \theta) - \langle\kappa\rangle(\theta) \tag{1.6.5}$$

The mean tangential shear at radius θ is therefore a direct measure of the total projected mass inside this radius, minus a boundary term (Schneider, P. et al, 2006). Therefore, galaxy-galaxy lensing is a probe of both the dark and luminous matter contained around a galaxy. 'weighing' galaxies using this technique over time can inform us of the growth of structure over time, which depends on the expansion of the universe.

1.7 Cosmic Shear

It is convenient to write the complex shear as $\gamma = |\gamma|e^{-2i\phi}$. Then we may define the tangential and cross components of the shear by taking the real and imaginary parts of the shear, respectively (Mandelbaum, 2018)

$$\gamma_+ = -\mathrm{Real}[|\gamma|e^{-2i\phi}] \;\; \gamma_x = -\mathrm{Imag}[|\gamma|e^{-2i\phi}] \tag{1.7.1}$$

We can then define two two–point correlation functions that define cosmic shear

$$\xi_+ = \langle \gamma\gamma^* \rangle = \langle \gamma_t\gamma_t \rangle + \langle \gamma_x\gamma_x \rangle \tag{1.7.2}$$

$$\xi_- = \mathrm{Real}[\langle \gamma\gamma \rangle e^{-4i\phi}] = \langle \gamma_t\gamma_t \rangle - \langle \gamma_x\gamma_x \rangle \tag{1.7.3}$$

1.8 Intrinsic Alignments

Intrinsic alignments (IA) are the coherent correlation of galaxy shapes in excess of any weak lensing signal due to physical interactions, gravitational interactions between galaxies. Halos of elliptical galaxies in close physical proximity to each other are stretched coherently by a tidal gravitational field. Because of IA, we may no longer assume that the intrinsic ellipticities of galaxies are random in orientation. This means IA is a systematic source of uncertainty in lensing analyses. In particular, there are three terms of interest that contaminate the shear-shear signal (Mandelbaum, 2018).

$$\langle \epsilon_i^s \epsilon_j^{s*} \rangle; \langle \epsilon_i^s \gamma_j^* \rangle; \langle \gamma_i \epsilon_j^{s*} \rangle \tag{1.8.1}$$

The first term, *II*, is the correlation between the intrinsic ellipticities of the ith and jth galaxy. This is expected to be non-zero only if the galaxies are in close physical proximity to each other. The second and third terms, the *GI* terms are the correlation of the intrinsic ellipticity of a galaxy with the shear of another. These are non-zero when the foreground galaxy ellipticity has to be correlated via IA to structures that shear a background galaxy (Kilbinger, 2015).

1.9 Weak Lensing Observables

The real space correlation functions for cosmic shear are formally defined as a linear combination of $\xi_{tt}(\theta)$ and $\xi_{\times\times}(\theta)$, which are directly measured from galaxy shapes

$$\xi_{tt} = \frac{\sum_{i<j} w_i w_j e_{t,i} e_{t,j}}{\sum_{i<j} w_i w_j} \tag{1.9.1}$$

and

$$\xi_{\times\times} = \frac{\sum_{i<j} w_i w_j e_{\times,i} e_{\times,j}}{\sum_{i<j} w_i w_j} \tag{1.9.2}$$

where $e_{t/\times,i}$ is the tangential/cross component of the complex ellipticity of the ith galaxy. We also use a weighting scheme:

$$w_i = \frac{1}{\sigma_{SN}^2 + \delta e_i^2} \tag{1.9.3}$$

here σ_{SN} is the shape noise, typically $\approx .35$ and δe is the error in ellipticity measurement. Finally, we define

$$\xi_+ = \xi_{tt} + \xi_{\times\times} \tag{1.9.4}$$

and

$$\xi_- = \xi_{tt} - \xi_{\times\times}$$

(1.9.5)

1.10 The Work Ahead

Thus far we have covered some cosmology fundamentals, and introduced three types of two-point correlation functions that can probe the matter field, and some potential sources of systematic error. In the following chapter, we will apply these probes to real observational data in an attempt to constrain cosmological parameters. In doing so, we will address a new source of systematic error that has not yet been considered for experiments of this nature. Until now, the leading systematic on the predicted redshift distribution of galaxies has been getting an accurate mean of the distribution. In the work presented here, we will address the degree to which the *shapes* of the predicted redshift distribution of galaxies affect the cosmology constraints that are ultimately obtained.

In the next chapter thereafter, we will return our attention to predicted redshift distributions of galaxies, and examine how they are impacted by galaxy blending. This systematic error on predicting redshift distributions, and it's subsequent effect on obtaining cosmology constraints has only begun to be investigated in the literature. In our work here, we statistically characterize errors in redshift estimation due to galaxy blending in realistic image simulations, and compare them to the allowed tolerances for two point correlations in upcoming so called "Stage IV" experiments.

Chapter 2

Cosmology Constraints From Joint Probes

2.1 Introduction

The large scale structure (LSS) today is imprinted with the initial matter power spectrum, and by extension, the initial conditions of the universe, billions of years ago. The statistical properties of the LSS and their time evolution are governed by the cosmic acceleration, matter-energy content, and gravity. Cosmology surveys leverage this sensitivity as a means to understand the fundamental nature of dark matter and dark energy. A prerequisite for many of these studies is knowledge of the redshift of observed galaxies. Surveys rely increasingly on photo-zs to satisfy this requirement. However, as data sets become photometrically deeper, and new probes are introduced, photo-z's are becoming a leading systematic error in cosmology inference.

Two point statistics of photometrically observed galaxies have been recognized as powerful techniques for characterizing the LSS. Galaxies are divided into redshift bins, which enables measurements of time evolution of the LSS (Albrecht, et al., 2006; Huterer, et al., 2015; Zhan & Tyson, 2018). This requires sufficiently accurate redshifts for every galaxy used. Obtaining a complete spectroscopic redshift sample to complement the photometric data is intractable. Photometric samples are dominated by many millions of faint galaxies, for which obtaining high signal to noise spectroscopic redshifts (spec-zs) is prohibitively costly. Instead, photo-z algorithms are combined with much smaller samples of spec-zs to generate redshift estimates. The spectroscopic samples are often photometrically non-representative of the galaxies for which photo-zs are measured, and re-weighted to correct for this.

photo-zs have been used in past (Jee et al., 2016; Kilbinger, et al., 2013), and present (Hikage et al., 2018; Hildebrandt et al., 2018, 2017; Troxel et al., 2018) weak lensing surveys, and are planed for future surveys (Euclid, WFIRST, LSST). These experiments measure statistical properties of shapes of distant galaxies. Light emitted from these distant galaxies suffers a series of deflections due to the gravitational field of the LSS it encounters. The observational signature, *cosmic shear*, of this effect is the coherent alignment of galaxies as a function of angular separation on the sky. Cosmic shear primarily probes the matter field, and is consequently most sensitive to Ω_m and σ_8, the root mean square of mass fluctuations. Sensitivity to these parameters is degenerate; they are often combined to define $S_8 = \sigma_8(\Omega_m/.3)^\alpha$ to break the strong correlation between Ω_m and σ_8 in these experiments, where α is optimally selected to break the degeneracy.

The combination of cosmic shear with galaxy clustering and galaxy-galaxy lensing from the same survey data can reduce errors on S_8. These two additional probes both target environments around luminous galaxies. Because of the galaxy-halo connection, these probes exclusively observe over dense environments. As a result, these probes provide an enhanced measure of structure if used independently. We model this enhancement with a linear factor, b, the galaxy bias (Kaiser, 1984). If galaxy-galaxy lensing and galaxy clustering are used together in a joint analysis, however, one can solve for the linear galaxy bias and obtain an unbiased measure of structure formation.

Galaxy-galaxy lensing is the correlation of the shape of background galaxies with the position of foreground galaxies. The mass from the foreground objects act as a magnifying glass, distorting light from sources behind them into a coherent tangential alignment. Galaxy-galaxy lensing is linear in the galaxy bias, b, as the position of luminous foreground objects are targeted.

Galaxy clustering is the correlation between the galaxy positions. Because of the galaxy-halo connection, and because galaxies reside in the same LSS, their positions have non-zero correlation. Galaxy clustering is quadratic in the galaxy bias, as it correlates the positions of two galaxies.

The joint application of all three two point statistics is colloquially called the "3x2 point" analysis, and has been applied in previous experiments (Abbott et al. (2018); van Uitert et al. (2018)). An important feature of the 3x2 point analysis is that the joint combination of the individual cosmological probes can suppress errors individual probes suffer from (Zhan & Tyson, 2018). This benefit comes with a cost, since the additional probes are accompanied by additional systematic errors. Typically, one uses a cosmological model that treats the galaxy bias and error in mean redshift for each tomographic bin as nuisance parameters marginalizes over them. However, systematic sources of error that are degenerate with galaxy bias and mean redshift error which are omitted from the model introduce degenerate solutions. Specifically, errors in higher order moments of photo-z distributions (width and skew) can imitate the galaxy bias. In a scenario where errors in photo-z widths or skew are unaccounted for, the imprint they leave on measured two point statistics can erroneously be fit by an evolving galaxy bias. If left unaccounted for, such degeneracies have the potential to affect the cosmological parameter estimation.

The need for control in systematics comes into sharper focus when we consider the results of cosmic shear surveys alongside other cosmology experiments. A complete cosmological model must be able to accurately describe the universe across epochs. It is therefore interesting to compare constraints resulting from measurements of the Cosmic Microwave Background (CMB) at $z \sim 1100$ to dark matter and dark energy probes from $z \sim 1.5$ and lower. S_8 predictions from *Planck*, the highest precision CMB experiment to date, are 2-3σ above those of several cosmic shear results.

This S_8 tension may be providing a hint of new physics beyond the standard ΛCDM cosmological model. However, interpretation of these 2-3 σ tensions as a departure from the concordance model must be carefully examined. Weak lensing surveys suffer from a myriad of pernicious systematic effects, and the S_8 tension may be due entirely to these. Therefore, high precision measurements demand robust mitigation of systematic errors as a necessary prerequisite if they are to address the S_8 tension, or test cosmological models more broadly.

Control and mitigation of systeamtics in weak lensing surveys-on many fronts-has matured in recent decades. For example, advancements in the design of observational facilities, observation techniques, galaxy shape measurement algorithms, point spread function (PSF) measurement and modeling, and CCD physics have enabled high precision cosmology results in completed Stage II and ongoing Stage III experiments (Mandelbaum, 2018). However, treatment of photo-z errors and systematics have, in general, not matured at the same scale. Furthermore, photo-z performance may have been acceptable for previous and current cosmic shear results, but the improving quality of data, and combination of cosmic shear with other two point statistics of the LSS, create much more stringent requirements for photo-z performance. This is especially true for experiments designed to test the time evolution of dark energy and alternate dark energy models more broadly, which require systematic errors to be controlled at the sub-percent level.

In this work, we investigate the impact that realistic errors in the estimated redshift distribution of galaxies, the n(z), in a weak lensing experiment can have on cosmological inference from the 3x2 point technique. Stress testing n(z) errors on real data of similar quality to ongoing and future experiments can probe failure modes and error regimes that may not present themselves in simulations, or in an analytic treatment.

We utilize an LSST precursor experiment, the Deep Lens Survey, as a setting to evaluate the degree to which n(z) and galaxy bias estimation errors influence cosmology inference. While error effects due to catastrophic outlier rate, and error on n(z) mean have been considered in several studies, here we explicitly examine the extent to which degeneracies in n(z) width have on galaxy bias estimation, and cosmology inference subsequently.

2.2 Statistical Probes of Cosmological Structures

The probes used in this analysis form auto or cross correlations of tomographic bins of the projected angular galaxy field and the projected angular convergence field. We assume the galaxy field, δ_g, is a biased representation of the matter field, δ_m, and we may capture this bias with a linear correction: $b\delta_m = \delta_g$, where b is the galaxy bias. The galaxy field is sampled by observing galaxies with redshift distributions, $p(z) = p(\chi)(d\chi/dz)$ where χ is the line of sight comoving distance. We adopt the notation of using $p_g^i(\chi)$ and $p_\kappa^i(\chi)$ to represent the normalized redshift distribution of the ith tomographic bin of the galaxy and convergence field, respectively. Because the non-linear matter power spectrum, $P(k,z)$ gives rise to these fields, measures of them constrain $P(k,z)$, and cosmology by extension.

Under the flat sky Limber approximation, the Fourier space wave vector $k \to \frac{l+1/2}{f(\chi)}$ and $P(k,z) \to P\left[\frac{l+1/2}{f(\chi)}, z\right]$ where χ is the line of sight comoving distance and $f(\chi)$ is the comoving angular diameter distance.

The lth multipole moment of the shear-shear power spectrum is given by

$$C_{\kappa\kappa}^{ij}(l) = \int_0^{\chi_h} \frac{q^i(\chi)q^j(\chi)}{a^2(\chi)} P\left[\frac{l+1/2}{f(\chi)}, \chi\right] \mathrm{d}\chi \tag{2.2.1}$$

Here, χ_h is the line of sight comoving horizon distance, a is the scale factor, and q^i is the lensing efficiency of the ith tomographic bin, given by

$$q^i(\chi) = \frac{3H_0^2\Omega_m}{2c^2}\int_{\chi}^{\chi_h} p_\kappa^i(\chi')\frac{f(\chi'-\chi)}{f(\chi')}\,\mathrm{d}\chi' \tag{2.2.2}$$

where $p_\kappa^i(\chi)$ is the normalized redshift distribution of galaxies in the ith tomographic bin of the convergence field, c is the speed of light, and H_0 is the Hubble Constant. The galaxy clustering power spectrum is given by

$$C_{gg}^{ij}(l) = b^i b^j \int_0^{\chi_h} \frac{p_g^i(\chi)p_g^j(\chi)}{f^2(\chi)} P\left[\frac{l+1/2}{f(\chi)},\chi\right]\mathrm{d}\chi \tag{2.2.3}$$

where $p_g^i(\chi)$ is the normalized redshift distribution of galaxies in the ith tomographic bin of the galaxy field, and b^i is the linear galaxy bias in the ith bin.

The galaxy-galaxy lensing power spectrum is given by

$$C_{g\kappa}^{ij}(l) = b^i \int_0^{\chi_h} \frac{q^i(\chi)p^j(\chi)}{a(\chi)f(\chi)} P\left[\frac{l+1/2}{f(\chi)},\chi\right]\mathrm{d}\chi \tag{2.2.4}$$

The linear bias b^i in the galaxy-galaxy lensing power spectra is assumed to be identical to that in the galaxy clustering spectra (e.g. $r^i = b_\times^i/b^i = 1$).

We emphasize that in equations 2.2.3 and 2.2.4, the power spectra amplitude depend on both the galaxy bias *and* the n(z) distribution of the galaxies being considered. The covariance in these quantities may have a detrimental effect on the accuracy of any fit to a cosmology model if not captured correctly. N(z) models are required as input for cosmology inference. Although it is true the means of their distributions are treated as a nuisance parameter and marginalized over, this approach may not capture the impact higher order moments in the n(z) inflict.

To help motivate this argument, we consider a toy example. Suppose the true n(z) of a tomographic bin is broader than the estimated n(z). We proceed, calculating a model $w(\theta)$ autocorrelation using the estimated n(z) as input. The model signal amplitude will rise as galaxy positions in narrower redshift slices have stronger correlations. This will have a down-stream effect in model fitting; The fitted galaxy bias will produce a smaller value than the true galaxy bias, as the predicted model is already overestimated. It is therefore crucial to obtain an accurate n(z) *shape* as well as an accurate mean. We will revisit this matter later in section 2.6.2, where we consider empirical methods to correct the n(z) shape.

In practice, the measured data will be in real space, which will require us to transform the power spectra. This is achieved by applying Hankel transformations to the power spectra. The angular two point correlation function, tangential shear, and cosmic shear respectively are given by

$$w(\theta) = \frac{1}{2\pi} \int J_0(l\theta) C_{gg}(l) l \, \mathrm{d}l \tag{2.2.5}$$

$$\gamma_t(\theta) = \frac{1}{2\pi} \int J_2(l\theta) C_{g\kappa}(l) l \, \mathrm{d}l \tag{2.2.6}$$

$$\xi_{+/-}(\theta) = \frac{1}{2\pi} \int J_{0/4}(l\theta) C_{\kappa\kappa}(l) l \, \mathrm{d}l \tag{2.2.7}$$

where $J_n(l\theta)$ is the nth order Bessel function of the first kind. For this analysis the real space correlation functions are measured with estimators. The measurement uses observed galaxies as tracers of the galaxy field and convergence field. A full discussion of the estimators follows in section 2.4.

2.3 Data

The Deep Lens Survey (DLS) is a weak lensing survey designed as a precursor to the Large Synoptic Survey Telescope (LSST). Broadband imaging was obtained in $BVRz'$ to enable photo-z estimation and shape measurement. As a result, it provides a staging ground to examine the impact photo-z estimation has on determination of galaxy bias and cosmology constraints.

The DLS consists of five 4 square degree fields, distributed across the sky with wide angular separation between them. As a weak lensing survey, the fields were chosen to avoid the Milky Way plane and bright low-z galaxies, and otherwise chosen without regard to known structures. The total area of the footprint was chosen such that deep photometric data could be achieved for shape measurement and photometric redshift estimation. The area was evenly split into five widely separated fields to avoid cosmic variance. Each field (F1-F5) was subdivided into a 3x3 grid of 9 sub-fields which were observed with a shift and stare dithering pattern. Observations for the R band were only carried out in the best seeing conditions (with a typical full width half max of 0.9

arcseconds), accumulating 18000s in exposure time. The R band exposure time enables a typical 5σ detection limit of $m_R = 27$ for point sources. Observations in the remaining filters were 12000s each, and the seeing conditions in which they were observed was prioritized in the following order: V, B, z'. Because priority was given to the R band in exposure time and seeing, detection and shape measurement are carried out on the R band coadded images (Wittman et al., 2002).

F1 and F2 lie in the northern hemisphere, and were observed with the Mayall 4-m Telescope with the MOSAIC I camera at Kitt Peak National Observatory (KPNO). The remaining sub-fields lie in the southern hemisphere, and were observed with the Blanco 4-m Telescope with the MOSAIC II camera at Cerro Tololo Inter-American Observatory (CTIO). The focal planes in both cameras have a 4x2 array of 2k x 4k CCDs with .25 arcsecond pixel scale at the center of the field of view (FOV), spanning a 35 square arcminute FOV (Wittman et al., 2002).

The initial photometry was carried out using `Source Extractor` MAG_AUTO outputs (Bertin & Arnouts, 1996). Photometric calibration was performed on magnitudes and colors using `COLORPRO` to correct for variations in seeing conditions between visits. A global linear least-squares (`Übercal`) approach in the style of the Sloan Digital Sky Survey (SDSS) (Padmanabhan et al., 2008) was implemented and subsequently mitigated spatial variations in photometry (Wittman et al., 2012). Galactic extinction was corrected using the Schlegel reddening maps (Schlegel et al., 1998).

2.3.1 Shape measurement

Seeing conditions from the atmosphere, dome, telescope optics, and detector which define the point spread function (PSF) impart ellipticities to measured sources. To recover the true ellipticities of galaxies, these effects must be removed. Removal of these signatures in the shape catalog used in this analysis is fully described in Jee et al. (2013) and we briefly summarize it below.

The `Stack-Fit` (S-Fit) technique from Jee et al. (2007) is used to remove PSF effects on the coadd level. For each individual exposure, for each CCD, the PSF is modeled by fitting to high signal to noise stars. The stars are decomposed into a linear combination of their principle components (eigen PSF). Jee et al. (2013) show that 20 such components are sufficient to achieve an accurate model. A 3rd order polynomial is used to interpolate between modeled stars to create a spatially varying PSF on each chip for each individual visit. To construct a PSF on the coadd, the principal components for each individual exposure are stacked. Shape measurement is then carried out on coadded images by fitting stacked-PSF-convolved elliptical Gaussians to galaxies to estimate their semi-major and minor axis and orientation angle.

We work in the weak lensing limit, where the ensemble average of galaxy ellipticities, $\langle\epsilon\rangle$, is an unbiased estimator of the reduced shear, g. Deviations from the true reduced shear, g_{true} and the observed reduced shear, g_{obs} are given by a first order correction $g_{true} = (1+m)g_{obs} + C$. Jee & Tyson (2011) use realistic image simulations of the DLS observing conditions to empirically determine the quantities $m_\gamma = (1+m)$ and C as a function of R band magnitude and find

$$m_\gamma = 6 \times 10^{-4}(m_R - 20)^{3.26} + 1.036 \tag{2.3.1}$$

and C to be negligible. Similar to Choi et al. (2012), the sample of galaxies for which we use shapes is $22 \leq m_R \leq 24.5$, which corresponds to $1.04 \leq m_\gamma \leq 1.12$.

2.3.2 Photometric redshift estimation

The publicly available Spectral Energy Distribution (SED) template based code Bayesian Photometric Redshift (BPZ) (Coe et al., 2006) was used for photometric redshift estimation. A detailed account is presented in Schmidt & Thorman (2013), but an abbreviated discussion follows. The six default SED templates packaged with BPZ were empirically adjusted based on overlapping spectroscopic redshifts obtained from the Smithsonian Hectospec Lensing Survey (SHeLS) (Geller et al., 2005). A type and magnitude dependent prior was fit in the manner of Benítez (2000), and relied on two sources of training data. The SHeLS data were used to fit the prior for galaxies $m_R \leq 21$. For fainter galaxies up to $m_i \leq 24$, spectroscopic redshift data from VIMOS-VLT Deep Survey are used for training. R band magnitudes were used for the magnitude dependence of the prior, and the template dependence was marginalized over. After the templates and priors have been adjusted, BPZ takes galaxy magnitudes and the DLS filter transmission curves as input to calculate a χ^2

$$\chi^2(T,z,a) = \sum_i \left[\frac{f_i - a f_{Ti}(z)}{\sigma_{f_i}}\right]^2 \tag{2.3.2}$$

where f_i is the observed flux in the ith filter with error σ_{f_i}, $f_{Ti}(z)$ is the predicted flux for a template (SED) T at redshift z after convolution with the system throughput, and a is a template

Table 2.1: Summary of cuts and descriptive statistics of galaxy populations in tomographic bins used for this analysis. Count is total number of galaxies in bin, $\langle z \rangle$ is the mean redshift of the bin, determined by integrating the stacked p(z) of the bin. L is for lens sample and S for source sample

bin	z edges	$\langle z \rangle$	m_r	count
L0	0.37 - 0.48	0.422	20 - 22	24288
L1	0.48 - 0.60	0.532	20 - 22	33918
L2	0.60 - 0.80	0.679	20 - 22	29100
S0	0.40 - 0.60	0.500	22 - 24.5	121491
S1	0.60 - 0.80	0.693	22 - 24.5	88708
S2	0.80 - 1.0	0.887	22 - 24.5	59845

normalization factor. χ^2 values are calculated on a grid for all template sets for redshifts in the range $0 \leq z \leq 5$, and span of normalization factors. The likelihood is then $\mathcal{L} \propto \exp(-\chi^2/2)$. After weighting by the prior, the template set is marginalized over to produce a probability density function (PDF) $p(z)$.

Validating the photo-z performance requires an independent sub sample for which spec-zs are known. The PRIsm MUlti-object Survey (PRIMUS) provides us with such a distinct validation sample, as it overlaps with F5. The PRIMUS sub-sample is photometrically complete to R = 22.8, and contains a sample of randomly observed fainter objects that is 30% complete with $22.8 \geq R \geq 24.0$. The DLS photo-zs achieve a typical scatter of $\sigma_{photo_z} = 0.06(1 + z_{spec-z})$ with an outlier rate of 4% beyond 0.2(1 + z). The magnitude range for which templates were empirically refined, priors were trained, and validation was performed places constraints on the magnitude range for which we have high confidence in photo-z fidelity. This will be expanded on in the following section where galaxy selection and tomographic bin assignment are explicitly defined.

We will rely on the PRIMUS sub-sample when we attempt to empirically correct the mean and variance of the n(z)'s in 2.6.2. The PRIMUS sample was not used to train the prior. Because PRIMUS spectroscopic redshifts were not used to train the prior and instead reserved for validation we will use it to ensure our corrections are indeed making the estimated n(z) more representative of the true n(z).

2.3.3 Tomographic Selection

At heart, our work is concerned with the impact unaddressed n(z) errors leave on cosmology estimation. The photo-zs used to estimate the n(z)s for our samples must therefore be clearly defined.

We discuss the selections used to define our sample here, paying mind to the measurements that will be made on these data.

In equations 2.2.1, 2.2.3 and 2.2.4 it is evident that the galaxy field and the convergence field are the ‘primary ingredients’ in the two point probes. To directly measure these quantities, we must partition observed data into a *lens* sample, where galaxy positions trace the galaxy field, and a *source* sample, where galaxy shapes trace the convergence field. Below we discuss how these two samples are defined, and subsequently divided into tomographic bins. To begin, we discuss criteria that all galaxies used in this analysis must satisfy. Summary statistics that describe the tomographic bins are shown in Table 2.2.

Morrison et al. (2012) notes the 4 band filter set used for the DLS limits the reliable photo-z range from $0.4 \lesssim z \lesssim 1.0$. Consequently, our tomographic bins fall approximately in this range. To cull galaxies with unreliable photo-zs from the sample, we require galaxies are detected in all 4 photometric bands; The faintest galaxies permitted have $m_R \leq 24.5$, to allow sufficient signal to noise for robust photo-zs; the brightest galaxies allowed are $m_R \geq 20$ to exclude sources bright enough to saturate or have non-linear responses on CCDs. The bright cut also protects against bright M stars, a contaminant which can masquerade as red galaxies in the $0 \leq z \leq 1$ redshift range. The magnitude range for the experimental samples is representative of the magnitude range used to empirically adjust SED templates, train the prior, and validate photo-z performance as discussed in 2.3.2.

It is typical in tomographic analysis to use a point estimate-mode, mean, or median of the p(z) to assign galaxies to tomographic bins. This can be problematic, as it allows for PDFs which are multi-modal, overly broad, or otherwise unreliable. To filter for more reliable photo-z PDFS, we adopt the following technique. For a tomographic bin with edges z_l and z_h, we integrate each galaxy's PDF in the interval $z_l - z_h$. If the integrated area is greater than 0.5, its entire p(z) is added to the tomographic bin. All such galaxy PDFs which satisfy this criteria are subsequently stacked to form an estimated n(z) for that tomographic bin.

Lens sample: The magnitudes for all bins span $20 \leq m_R < 22$. To accommodate a source bin at higher redshift than the entire sample for the lens sample, we choose the total redshift range of $0.37 \leq z \leq 0.8$. This range is then subdivided into three tomographic bins. The width of the tomographic lens bins is largely set by the overall shape of the n(z); the galaxy counts at high redshift drop off rapidly after $z \sim 0.6$. We seek some empirical guidance on how to partition the bins. We calculate the cumulative distribution function (CDF) of a BPZ prior, which is calculated using the magnitude distribution for galaxies with R band magnitudes from 20 to 22. We then partition the CDF in the range $0.37 \leq z \leq 0.8$ into three roughly equal bins. This ultimately yields lens bins with redshift edges (0.37 - 0.48) , (0.48 - 0.6), (0.6 - 0.8).

Luminous red galaxies (LRGs) are often used to define a lens sample when calculating galaxy-galaxy lensing and galaxy clustering. This has several advantages: 1) the high mass of these objects creates higher signal to noise tangential shear measurements, given the same source plane. 2) red eliptical galaxies separate themselves from other galaxies in color color space as their SEDs are redshifted. This makes them ideal targets for template fitting photo-z codes, which can deliver accurate redshift estimates. 3) the galaxy bias is clearly defined for a particular galaxy sample (Rozo, et al., 2016).

Exclusively using LRGs delivers results which pertain to very specific parts of the matter density field, however. Aside from the possibility of introducing a bias, it is scientifically compelling to use more egalitarian lensing selections to probe the dark matter field more generally. From equation 2.2.4 and 2.2.3 we can see the power spectrum for galaxy clustering and galaxy-galaxy lensing is defined by the normalized redshift distribution of galaxies and the galaxy bias. This provides freedom to define less restrictive lens cuts, provided the n(z) can be estimated reliably.

The bright, low redshift, luminous galaxies in the DLS provide such a sample, which we use for our lens selection. These can provide a high shear signal for a more representative sample of galaxies in terms of e.g. their type and color. Consequently, a broader, more generalized statement about matter fluctuations follows from measurements on these galaxies.

Source sample: The magnitudes for all bins span $22 \leq m_R < 24.5$; which ensures that any one galaxy can only be placed in one lens bin or one source bin, exclusively. To allow these galaxies to be lensed by the foreground galaxies in the lens sample, we choose three tomographic bins with redshift intervals (0.4 - 0.6), (0.6 - 0.8), (0.8 - 1.0). Accurate shapes are imperative for shear estimation. Thus, a prerequisite for source galaxy is that the elliptical Gaussian fitting procedure used to measure its shape successfully converged. We then apply shape cuts to ensure the fitted ellipticities are reliable. The typical shape noise for the DLS is $\sigma_e \approx 0.25$, and so we require the shape error $\delta e < .25$, the typical amplitude of the shape noise, so it is comparable to-and never dominates-the

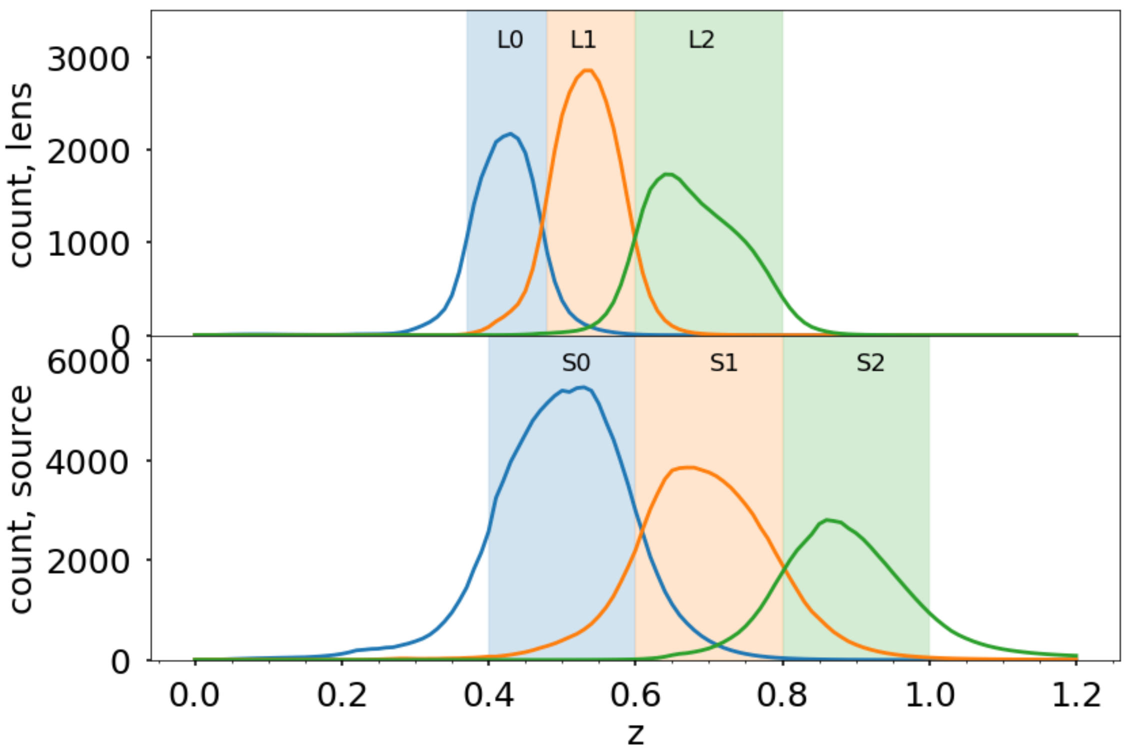

Figure 2.3.1: Redshift distributions for tomographic bins for the lens (top panel) and source (bottom panel) galaxies. Each solid curve shows the n(z) of a particular tomographic bin. The solid vertical colors indicate the redshift ranges of the tomographic bins of their corresponding colored n(z) curves. n(z)s for each bin are constructed by stacking the p(z) distributions for galaxies which have 50% of their area between the redshift interval of the bin.

shape noise. In Figure 14 of Jee et al. (2013), the authors demonstrate strong ellipticity bias in small galaxies-regardless of signal to noise and shape error. This shape bias is primarily from under sampling and pixelization effects, and can be avoided by requiring the semi-minor axis b measured from the elliptical Gaussian fitting procedure is larger than 0.4, which we enforce on all source galaxies. This cut protects against bias in preferred orientation angles in source galaxies.

In Figure 2.3.1 we show the resulting n(z) for both the lens sample (top panel) and source sample (bottom panel). n(z) distributions for L0, L1, and L2 are shown as the blue, orange and green curves in the top panel respectively, and the extent of these bins are indicated by the shaded regions of the same color. Similarly, the n(z) distributions for S0, S1 and S2 are shown in the bottom panel as the blue, orange, and green curves respectively, and the shaded regions indicate the width of the bins or corresponding color.

The various n(z) shapes for our tomographic bins are complex and influenced by the overall n(z) of the entire galaxy sample from the DLS. Selection criteria (integrated area inside of a bin, magnitude and shape cuts) additionally influence the shape of these curves. In particular these distributions have non-zero skew. We will only attempt to ameliorate errors to the first two moments in the n(z) distributions in this work, leaving higher order moments to future work.

One can gain intuition for the expected galaxy-galaxy clustering signal by examining the overlap in the n(z) of the lens samples. For example, L0 and L2 have a near zero over lap, and equation 2.2.3 suggests this will lead to vanishingly small amplitude for their cross correlation. Typically, the missmatch between the observed correlation and theoretically calculated correlation is explained by uniformly shifting the n(z) around or adjusting the galaxy bias. However, it is evident that the n(z) width-and overlap with neighboring n(z)-also contributes to the clustering signal amplitude. The degeneracy between changing the n(z) mean and n(z) width, and impact they have on cosmology inference is examined in greater detail in section 2.3.1.

2.4 Measured Two Point Statistics

Previously we have provided a theoretical definition of our cosmological probes, using the matter density power spectrum, n(z) distributions, and gravitational lensing effects. In practice, we use statistical estimators applied to observed data to obtain measurements. In this section we define the estimators used in this work for galaxy clustering and galaxy-galaxy lensing.

2.4.1 Galaxy clustering

If the likelihood of finding a galaxy inside a small patch on the sky subtended by solid angle Ω, is P_Ω, the likelihood of finding a neighboring galaxy an angular distance θ away is given by $P_\Omega^2[1+w(\theta)]$. $w(\theta)$ is the probability in excess of random correlations, and is expected to be non-zero when measured on galaxies which live in the same large scale structure. To measure $w(\theta)$ between two fields i and j we use the Landy-Szalay estimator (Landy & Szalay, 1993)

$$w(\theta) = \frac{\langle D_i D_j \rangle + \langle R_i R_j \rangle - \langle R_i D_j \rangle - \langle D_j R_i \rangle}{\langle R_i R_j \rangle} \tag{2.4.1}$$

where $D_i D_j$ is the cross correlation between data in the ith and jth fields, $R_i R_j$ is the cross correlation of the ith and jth random fields, and $D_{i/j} R_{j/i}$ is the cross correlation between the i/jth Data field and j/ith random field. For every DLS field, for every galaxy field tomographic bin, we generate 7 times as many randoms as data points. The case where $i = j$ yields auto-correlations. Most of the cosmological information is encoded in the auto-correlations, and the auto-correlations are exclusively used to constrain cosmological models.

$w(\theta)$ is sensitive to masking effects, and having accurate masks and boundary regions which match the data and simulated random points is crucial. For our lens sample and generated randoms, we use the masks provided courtesy of the DLS. The masks will be used again later in section 2.7 when we generate many realizations of the DLS, to ensure they adhere to the true survey geometry. As Morrison et al. (2012) indicates, the DLS has low signal to noise regions between subfields and along the edges of the field footprints due to the dithering pattern used. The lens sample defined for this work, however, is comparatively bright, going as faint as $m_R = 22$. At this magnitude range, the survey is depth complete, even in the lower signal to noise regions. As Yoon et al. (2019) point out, for such a bright magnitude cut, we may neglect higher order corrections due to lower signal to noise, extinction and so forth.

Because of the size of the DLS fields, our measurements of $w(\theta)$ rely on galaxies in localized sky regions, namely each of our fields F1-F5. $w(\theta)$ measurements, consequently, will be biased with respect to the true, global, value of $w(\theta)$. This deviation can be corrected by adding the integral constraint. The additional offset is typically calculated by minimizing the chi square of a power law model fit to the measured $w(\theta)$ signal (Yoon et al., 2019), where the initial guess of model parameters are motivated by a cosmological model. Matthews & Newman (2012) noted that angular bins for $w(\theta)$ are known to be correlated, and consequently knowing the co-variance of $w(\theta)$ is necessary to obtain an accurate model fit. A complete discussion of the full covariance matrix-including the covariance of $w(\theta)$ measurements-of both two-point statistics is presented later in section 2.7. However, we will note here for clarity that in essence the covariance matrix is estimated by calculating the covariance of many measurements of $w(\theta)$ on simulated DLS realizations.

The fitting function is of the form $w(\theta) = A(\theta)^{1-\gamma} - C$ where C is the integral constraint, A is a constant, and γ parameterizes the power law slope. Matthews & Newman (2012) also indicate that fitting all three parameters at once is highly degenerate. To obtain reliable fits for all three parameters, we fix γ and fit for A and C by reducing the χ square. We then repeat this procedure for a vector of γ values, and select the combination of parameters for which the χ^2 is smallest.

The measured $w(\theta)$ signal for all unique pairs of lens bins is shown in individual panels in Figure 2.4.1 as blue data points. The top right corner in each panel contains a pair of integers, which designates the lens bins being correlated (e.g. 00 auto correlates L0, and 21 cross correlates L2 and L1). The error bars on the data points come from the covariance matrix, which is estimated by realizing the DLS many times using simulations, the full details of which are discussed in section 2.7. Each panel also contains a solid line, which is the theoretical prediction given the best fit cosmology. Finding the best fit cosmology is fully explained in 2.7.

2.4.2 Galaxy-galaxy lensing

The gravitational field of massive foreground galaxies distorts the shapes of background galaxies, making them coherently tangentially aligned. Thus, measuring the shapes of background galaxies informs us of the matter distribution that host the foreground galaxies.

We first formally define the tangential and cross components of the ellipticities of background sources.

$$e_t = -\Re[e \exp(-2i\phi)],\ e_\times = -\Im[e \exp(-2i\phi)], \tag{2.4.2}$$

where e is the magnitude of the complex ellipticity, and ϕ is the azimuthal rotation angle which

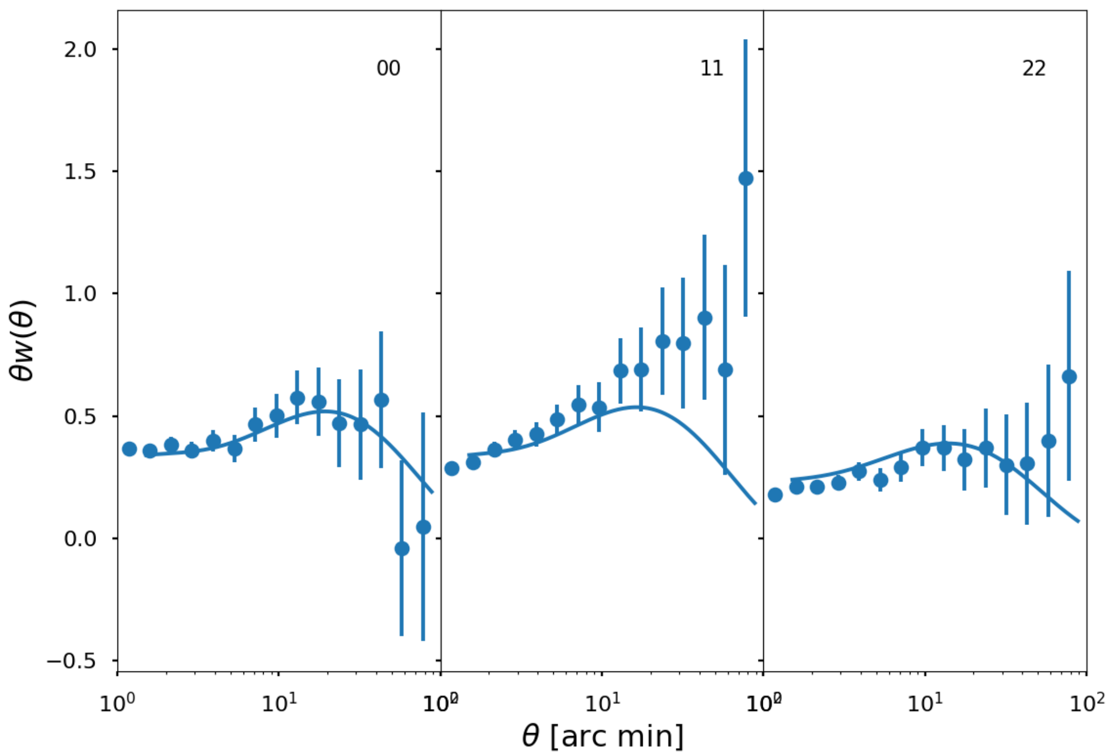

Figure 2.4.1: Measured Galaxy clustering auto and cross correlation function for three tomographic bins from the lens sample, (0.37 - 0.48), (0.48 - 0.6), (0.6 - 0.8) in blue points. The solid lines show the theory prediction of the best fit cosmology from our cosmology inference. Number pairs in top right corner of the panels indicate which bins are being correlated, e.g. 00 is the auto correlation of the (0.37 - 0.48) bin. Because the points and errors are correlated, we caution the reader from doing a "chi-by-eye" fit. We address the high amplitude of the last point in the 11 correlation later in section 2.9

defines the source galaxy's position relative to the lens galaxy it is being correlated with. For an individual foreground lens, the galaxy-galaxy lensing signal is

$$\gamma_{t/\times}(\theta) = \frac{\sum e_{i,t/\times} w_i}{\sum w_i} \tag{2.4.3}$$

where $e_{i,t/\times}$ is the tangential/cross component of the complex ellipticity of the ith source galaxy from a tomographic bin, and w_i is its weight. We define the per-galaxy weight as the inverse of the quadrature sum of the shape noise and shape error. True astrophysical signals are expected to have zero cross component signal absent the effects of source clustering and multiple lensing in cluster environments (see e.g. Bradshaw, Jee & Tyson (2019)). The amplitude of the cross signal can, however, be used for a null test for systematic errors, and motivate scale cuts. The measurement and discussion of the cross signal will be discussed later in this section.

We then stack up the tangential and cross signal across galaxies in each lens-source tomographic bin combination. Finally, a field geometry correction is applied by subtracting the tangential signal around randomly placed points (Singh, Mandelbaum, Seljak, Slosar & Vazquez Gonzalez, 2017). This ultimately makes our galaxy-galaxy lensing estimator

$$\langle\gamma_{t/\times}\rangle = \langle\gamma_{t/\times}(\theta)_g\rangle - \langle\gamma_{t/\times}(\theta)_r\rangle \tag{2.4.4}$$

where g (r) is for the galaxy (random) field.

The measured tangential shear are shown in Figure 2.4.2 as blue points. The top right corner of each panel contains two integers which indicate the lens and source bins are being correlated together (e.g. 00 cross correlates galaxy positions of L0 with galaxy shapes in S0, and 12 cross correlates L1 positions with shapes from S2. Note, when the first integer is larger than the second, the source population is at lower redshift than the lens population).

2.4.3 Shear-shear

Light from observed, distant galaxies is coherently distorted as it is lensed by the intervening LSS. Unlike galaxy-galaxy lensing and galaxy clustering, this two point correlation between galaxy shapes does not suffer from the galaxy bias. Galaxy-galaxy lensing and galaxy clustering exclusively probe

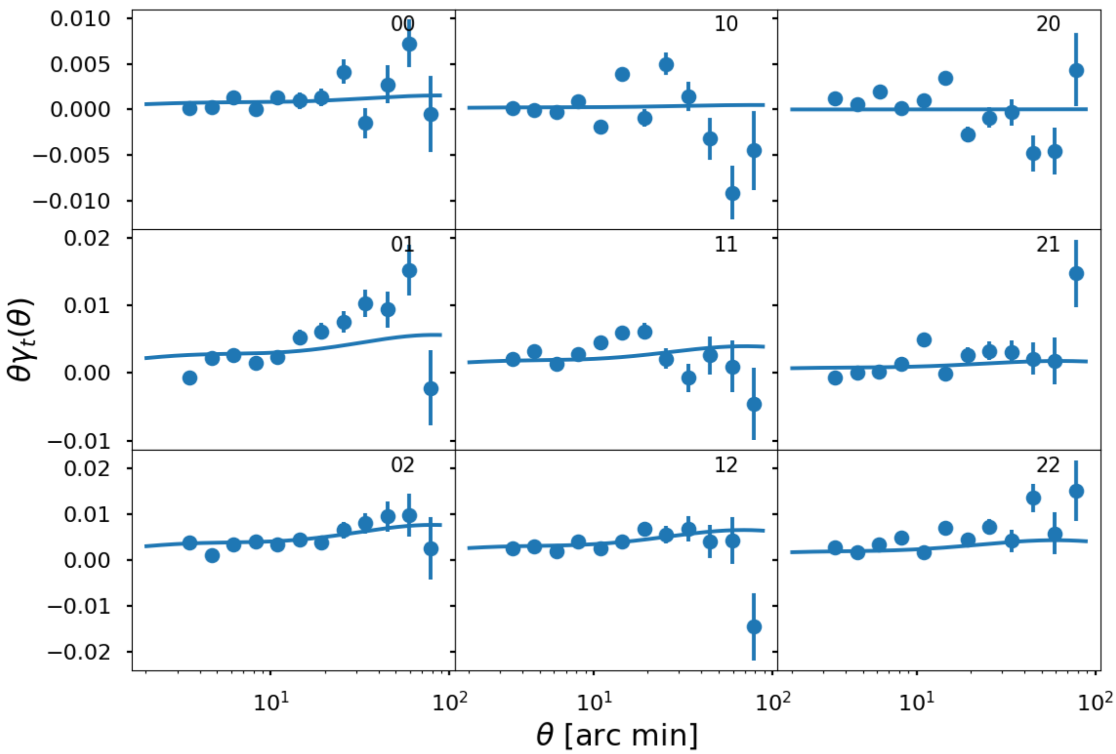

Figure 2.4.2: Galaxy-galaxy lensing signal (blue points) where positions from the lens sample are correlate with shapes from the source sample. The solid lines show the theory prediction using the best fit cosmology from our cosmology inference. Numbers in the upper right indicate which lens bin (left number) is correlated with which source bin (right number). Again, we caution the reader from doing a "chi-by-eye" fit since the data and errors are correlated.

environments around dense, massive, collapsed dark matter halos. Cosmic shear, on the other hand, is even-handed in its sensitivity to the matter field. Over and under densities both produce observational signatures.

An extensive discussion of cosmic shear measurements in the DLS can be found in Jee et al. (2013) and Jee et al. (2016), where the first cosmology constraints from the DLS were published. We offer an abridged discussion here. The real space correlation functions formally defined in equation 2.2.7, $\xi_+(\theta)$ and $\xi_-(\theta)$, are constructed with a linear combination of $\xi_{tt}(\theta)$ and $\xi_{\times\times}(\theta)$, which are directly measured from galaxy shapes

$$\xi_{tt} = \frac{\sum_{i<j} w_i w_j e_{t,i} e_{t,j}}{\sum_{i<j} w_i w_j} \tag{2.4.5}$$

and

$$\xi_{\times\times} = \frac{\sum_{i<j} w_i w_j e_{\times,i} e_{\times,j}}{\sum_{i<j} w_i w_j} \tag{2.4.6}$$

where $e_{t/\times,i}$ is the tangential/cross component of the complex ellipticity of the ith galaxy. The same weighting scheme for galaxy shapes used in galaxy-galaxy lensing is used here, where

$$w_i = \frac{1}{\sigma_{SN}^2 + \delta e_i^2} \tag{2.4.7}$$

here σ_{SN} is the shape noise, typically $\approx .35$ and δe is the error in ellipticity measurement. Finally, we define

$$\xi_+ = \xi_{tt} + \xi_{\times\times} \tag{2.4.8}$$

and

$$\xi_- = \xi_{tt} - \xi_{\times\times} \tag{2.4.9}$$

In Figures 2.4.3 and 2.4.4 we show the results for ξ_+ and ξ_- from shear-shear. The measured correlation function is shown in blue points in both cases, and the top right in each panel show which source tomographic bins are being correlated together. In solid blue lines, we show the results of best fit cosmology for a ΛCDM model. Unlike the previous correlation functions, there is a very noticeable discrepancy between the measured data and best fit cosmology. In particular, we see high scatter in both ξ_+ and ξ_-. It appears likely that the error bars are underestimated for the data. Taken together, it seems likely there are unaccounted for systematic errors in the data which are leading to the poor fit.

It is worth noting that we have a notably different approach to Jee et al. (2016) when measuring cosmic shear, in particular with regard to photometric redshifts requirements. We require detection in all four photometric bands, and we require that half of the area of each individual p(z) lies in the range of each tomographic bin. In Jee et al. (2016), only an R band detection is required, and only 11% of the intergrated p(z) is required to live in each tomographic bin. Additionally, the photometric cut is different between approaches. Here we restrict our sample to galaxies with an R band magnitude of 24.5 and brighter, while the previous study uses galaxies as faint as $m_R = 26$, being dominated by many fainter galaxies. The difference in magnitude cuts along with photo-z cuts means the samples differ considerably. One possibility for the discrepancy is that the shear calibration used in Jee et al. (2016) was appropriate for the photometric sample used in that study, but is inappropriate for the sample being considered in this study.

2.5 Pipeline Validation

In order to trust our results and demonstrate their robustness, it is incumbent upon us to demonstrate our analysis pipelines and methods can provide accurate results. Using observational data to achieve this end can be somewhat problematic, as strictly speaking we do not know the ground truth about the universe and its properties. Instead, we can perform pipeline tests on simulations, for which we know the input cosmology and other underlying properties exactly. To this end, we use a simulation to measure our two point statistics to verify that we are calculating them correctly on our DLS data set.

We use the Cosmological Data Challenge 2 (CosmoDC2) truth catalog. CosmoDC2 is a large synthetic galaxy catalog designed to emulate LSST data quality. This allows the scientific community to prepare for LSST first light and beyond by exploring survey specific systematics, like PSF estimation, and preparing analysis methods and pipelines ahead of time. The catalog is complete down to an r band magnitude of 28, occupies 440 deg^2 of sky area, with sources up to redshift $z = 3$. The catalogs provide a plethora of data quantities for each galaxy in the N-body simulation,

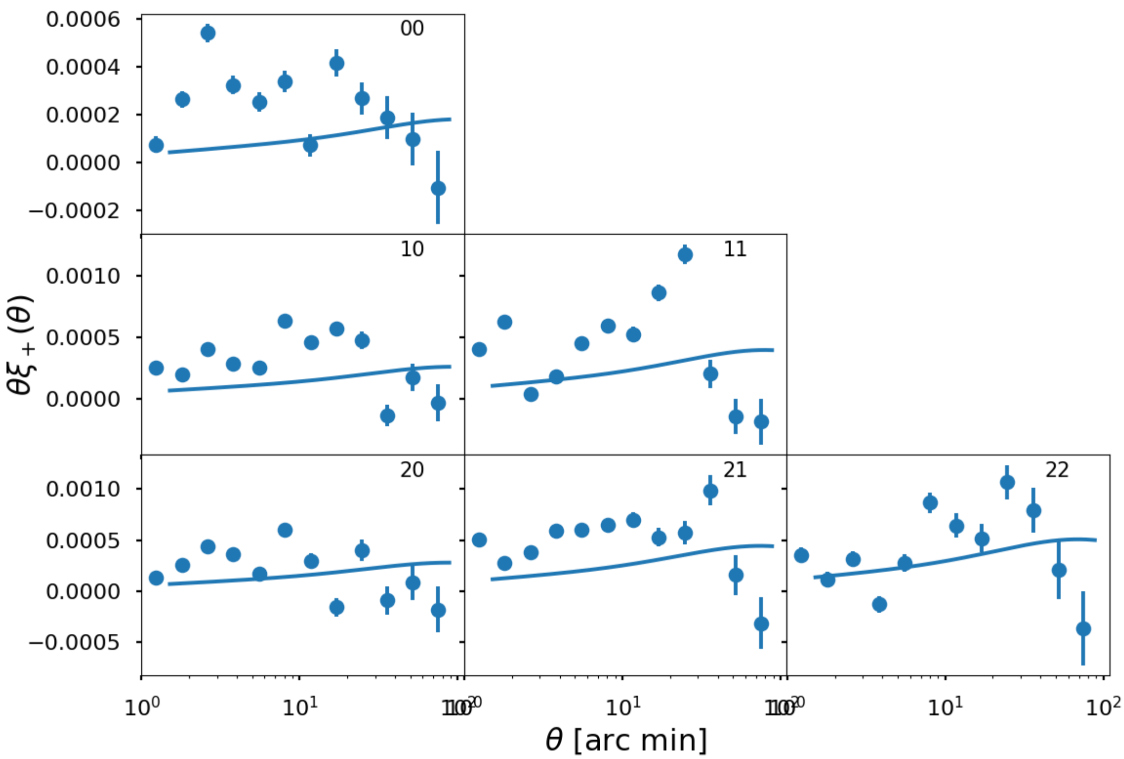

Figure 2.4.3: ξ_+ signal from the shear-shear correlation are shown in blue points, with best fit cosmology shown in solid blue lines. The top right corner of each panel shows which source bins are being correlated together. There is significant discrepancy between the measured theory and best fit cosmology, that cannot be explained by the errorbars. It is likely that the errors are underestimated, not accounting for unresolved systematic effects.

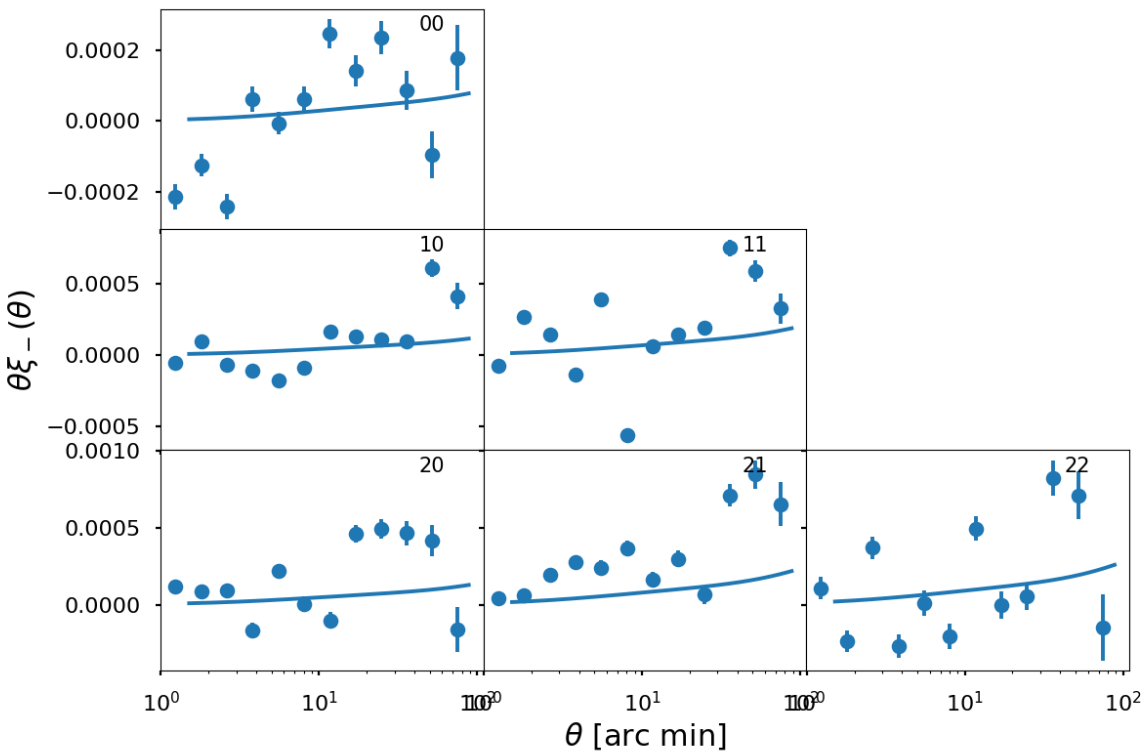

Figure 2.4.4: ξ_- signal from the shear-shear correlation are shown in blue points, with best fit cosmology shown in solid blue lines. The top right corner of each panel shows which source bins are being correlated together. Like, the ξ_+ signal, there is significant discrepancy between the measured theory and best fit cosmology, that cannot be explained by the errorbars. It is likely that the errors are underestimated, not accounting for unresolved systematic effects.

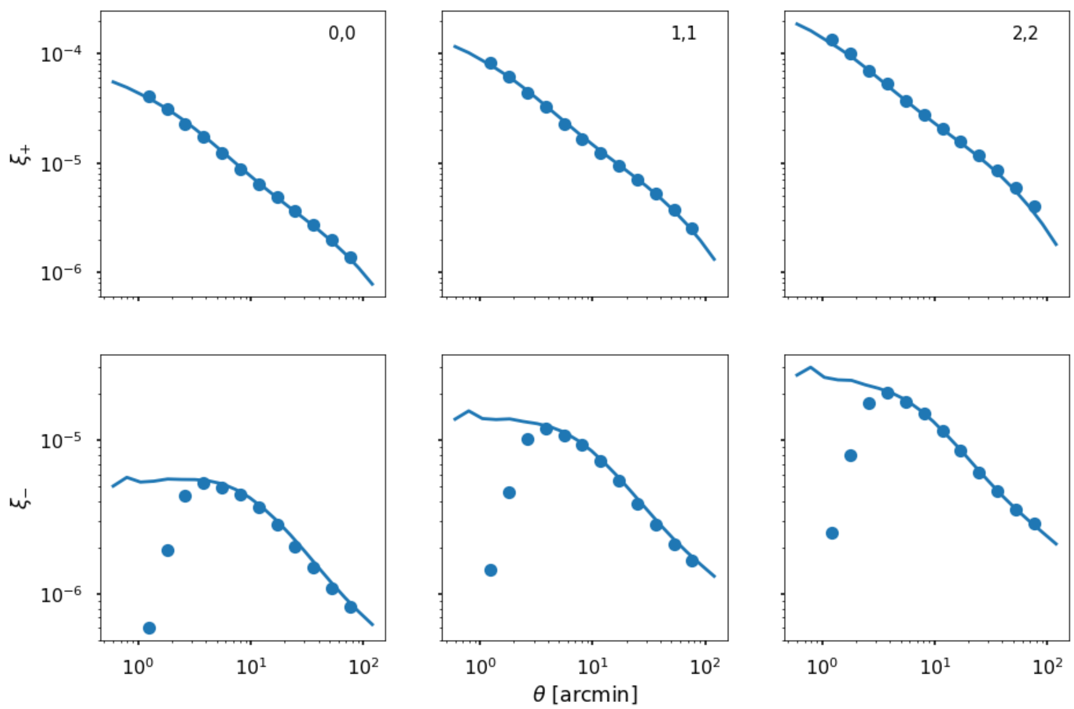

Figure 2.5.1: We show results of our two point calculation pipeline when ran on Cosmological Data Challenge 2, a truth catalog derived from an N body simulation. The top row shows measured ξ_+ in blue points for auto correlations for three tomographic bins. Theory predictions, using the input cosmology from the simulation, are shown as solid blue lines. The measurement and theory predictions show remarkable agreement. The bottom row shows measured ξ_- in blue dots and theory prediction in solid blue lines for the same three tomographic bins. The measured signal shows discrepancy with theory at the 1 - 10 arcminute range. This behavior is expected, due to the limiting resolution of the simulation used. Note, we did not include masks in the simulation, due to bright stars. The catalog used was an extragalactic catalog, and did not contain foreground Milkyway stars. Additionally we use true redshift distributions. Therefore, this measurement validates the shear measurement in isolation of masking and photo-z effects.

including broadband filter magnitudes, redshift, shape, convergence, among many others at the catalog level. Given the extensive testing and quantities provided by the cosmoDC2 catalog, it is ideally suited to our tests.

in Figure 2.5.1 we show shear-shear results of our validation run on CosmoDC2. The top panel shows measured ξ_+ signal in blue points, and the bottom panel shows measured ξ_- in blue points. Because the input cosmology and the n(z)'s of the tomographic bins are known exactly in this scenario, we can create theory predictions for these measurements. These are shown as solid blue lines in both top and bottom panels. The top panels show remarkable agreement between measured and theory prediction. The bottom panel shows good agreement at high angular separation, from 10 to 100 arcminutes. However, the measured points at 3 arcminutes and lower undershoot what is predicted by theory. This is expected due to resolution effects in the CosmoDC2 simulation at small angular separations, and not a fault of our pipeline (R. Mandelbaum, private communication, LSST Dark Energy Science Collaboration et al 2020).

2.6 Corrections to n(z) First And Second Moments

Validation p(z)s of a galaxy sample requires knowledge of the true n(z) of that sample. The SHELS and VVDS data do provide a sample of spec-zs, however they are used to tune template sets and train the prior for our BPZ run. Consequently, they cannot provide unbiased metrics for photo-z performance.

Instead, we use spec-zs from the PRIsm MUlti-object Survey (PRIMUS) (Coil et al., 2011), which were unused in any way to train the photo-zs. PRIMUS overlaps with 7 of the 9 subfields in F5 but does not cover the DLS footprint completely. Because SHELS and VVDS overlaped with F2, PRIMUS provides a distinct sample in a region on the sky widely separated from training galaxies. PRIMUS is depth complete to $R = 22.8$ and additionally observed 30% of galaxies with magnitudes $22.8 \geq R \geq 23.5$, which were chosen at random. Thus, the lens sample has complete coverage in terms of magnitude, while the source sample is incomplete in terms of depth and magnitude range. We also note that properties of galaxies in F5 (even galaxies that are depth complete) are not guaranteed to be fully reflective of galaxies in F1-F4.

2.6.1 Photo-z Mean

To assess the impact of error in the mean of the stacked photo-zs, we define the $z-$shift parameter, Δz, as the difference between the true n(z) mean and mean of stacked p(z). By translating the estimated n(z) by Δz

$$p^i(z) \rightarrow p^i(z + \Delta z_i) \tag{2.6.1}$$

—e.g. in the method of Yoon et al. (2019) & Hikage et al. (2018)-the estimated mean is made to match the true n(z) mean. We measure Δz_i for each tomographic bin, i for both lenses and sources by first stacking the PRIMUS-DLS sample p(z) in the method of 2.3.3. We then construct n(z)s using the corresponding spec-zs for each bin, and calculate the difference in means. The size of these shifts for the lens bins are 0.003, -0.003 and -0.039. For the source bins we have 0.017, 0.002, and -0.037.The mean differences obtained by this process will be reused in section 2.7, where we treat δz as a nussiance parameter to margnialize over.

2.6.2 Photo-z Width

In equation (2.2.3) the amplitude of the galaxy clustering power spectrum is not only affected by the linear galaxy biases; the overlap and shape of the n(z) of the tomographic bins being correlated contributes to the amplitude as well. In effect, an error in the n(z) width can imitate a cosmological signal that may erroneously attributed to the galaxy bias or n(z) mean. This is potentially problematic for cosmological inference: the joint combination of galaxy clustering and galaxy-galaxy lensing is meant to interlock so the bias is removed. However, if degeneracies arising from n(z) errors are unchecked, obtaining a high fidelity galaxy bias may not be achievable. It is therefore crucial to ensure the overlap in the wings of the n(z) is accurate. The ultimate impact on cosmology inference will be addressed in section 2.7.

The Probability Integral Transform (PIT) can be used to evaluate if the set of redshift PDFs for galaxies is a reasonable description of their true underlying n(z). Formally, the PIT for a single galaxy, for which we have knowledge of the true redshift, is defined as

$$\mathrm{PIT} = \int_{-\infty}^{z_t} p(z)\,\mathrm{d}z \tag{2.6.2}$$

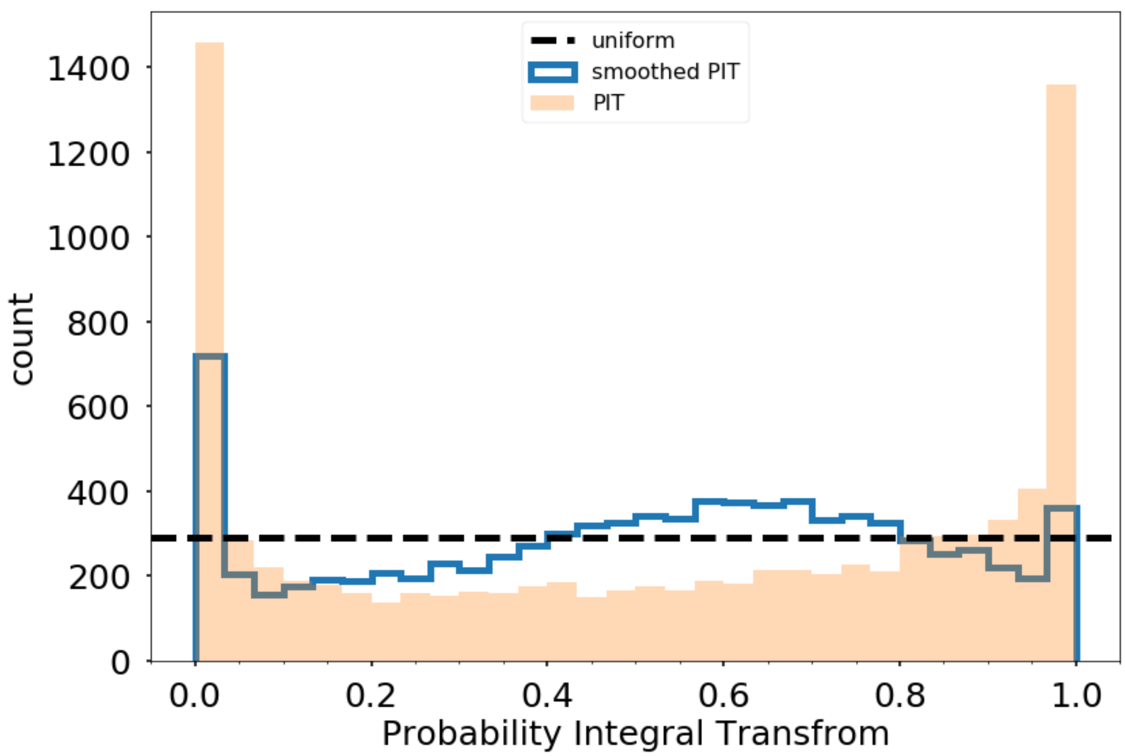

Figure 2.6.1: PIT histograms for galaxies in our sample for which we have an accompanying spec-zs from PRIMUS. For each galaxy in the DLS-PRIMUS sample, we calculate the cumulative distribution functions of their photo-z PDF up to the true redshift. If the ensemble of PDFs were an accurate description of the true underlying n(z), the histogram of all such values would be uniform (dashed black line). The orange distribution is the PIT histogram of DLS-PRIMUS galaxies before they are broadened with an optimal filter. The 'pile up' at the edges indicates the PDFs are overly narrow. The blue histogram shows the distribution after each individual PDF is convolved with a Gaussian. The width of the Gaussian kernel is calculated by splitting the galaxy sample into 4 roughly equal magnitude bins, and minimizing the Kullback-Leibler divergence for each magnitude bin. The optimal smoothing is able to broaden individual p(z)' so they form a more representative description of the true n(z) of galaxies considered. Regardless, a persistent bias is evident in the blue histogram as it trends upwards from left to right, indicating the right p(z) tails truncate faster than the left tails, suggesting the p(z)s may not capture the n(z).

where z_t is the true redshift and $p(z)$ is a galaxy's PDF. If the ensemble of PDFs are faithful descriptions of their analogous underlying n(z), the histogram of their PIT values is expected to be uniform. If the PDFs are overly broad, the PIT histogram is expected to peak near its center. On the other hand, if the PDFs are overly narrow or suffer from many catastrophic outliers, the PIT histogram will 'pile up' at the limits of the histogram range. Calculating the PIT requires knowledge of the true redshift, so we are restricted to calculating it for the sub sample of DLS data that have spec-zs from PRIMUS.

In Figure 2.6.1 we show the PIT histogram of the PRIMUS-DLS sample in solid orange. The build up at the limits of the PIT histogram is indicative of overly narrow PDFs. One possible explanation is that the photometric errors used in BPZ were underestimated; `Source Extractor` used for flux and flux error measurements in DLS processing, and is known to underestimate photometric errors redneed citation. Additionally, Wittman, Bhaskar & Tobin (2016) demonstrate that absence of uncertainty in BPZs SED templates lead to overconfident p(z)s. Under-predicted and unaccounted photometric errors subsequently yield overly confident individual PDFs. Following the method of Hoyle et al. (2018) and Wittman, Bhaskar & Tobin (2016), we posit that the widths of the stacked photo-z PDFs can be made more reflective of their n(z) by convolving their individual photo-z PDFs by a Gaussian filter

$$p(z) \rightarrow N(\sigma) \otimes p(z) \tag{2.6.3}$$

where $N(\sigma)$ is a Gaussian with width σ. To find an optimal width, we minimize the Kullback-Leibler divergence. For two discrete distributions, $P(x)$ and $Q(x)$, the Kullback-Leibler divergence is defined as

$$KLD = \sum_i P(x_i) \log[P(x_i)/Q(x_i)] \tag{2.6.4}$$

and quantifies the difference between the distributions. We let Q be a uniform distribution, and fit for a Gaussian width which minimizes the divergence when convolved with the photo-z PDF before the PIT is calculated. A Kullback-Leibler divergence value of 0 in this case would indicate the PIT histogram is identical to a uniform distribution.

We partition our spectra sample into 4 bins by R magnitude, where each bin has roughly the same number of spec-zs. The bins are (20 - 21.65), (21.65 - 22.27), (22.27 - 22.7), and (22.7 - 23.5). The optimal widths obtained for the bins are 0.045, 0.0695, 0.0896, and 0.128 respectively, and measure a Kullback-Leibler divergence of 0.053, a decisive improvement over the 0.276 before convolution. The resulting PIT histogram from broadening PDFs is shown in Figure 2.6.1 in the unfilled blue histogram. The post-convolution PIT histogram is much closer to the uniform distribution (dashed black line) than the pre-convolved PIT histogram (solid orange), consistent with our Kullback-Leibler divergence calculations.

Additionally, we train the widths for each individual tomographic bin. The per bin training yielded a Kullback-Leibler divergence of .0627. Aside from being sub-optimal to the magnitude training, this raises concerns about over training. A common technique to check for over training is to set aside part of the data for validation, and part for training. The dearth of galaxies with spec-zs in this sample makes this pursuit inefficient. Splitting the sample would produce lower fidelity fits in addition to giving noisy validation results. Thus, we proceed using filter widths from the magnitude training.

One caveat to this approach is that we are assuming the training sample used in F5 alone is completely representative of the galaxies in the other fields. The widths for bins in the F5 sample may be completely different than those in the other fields. However, since we do not have an accompanying spectroscopic data set for the other fields, we must rely on the PRIMUS data, and work under the assumption that it is indeed representative.

Once we have determined the optimal filtering width as a function of R band magnitude, we iterate over our PDFs, convolving them with their appropriate Gaussian filters according to the training model. The individual broadened PDFs are then assigned to tomographic bins such that their bin membership is identical to that before the PDFs were broadened. Finally, PDFs in the same bins are summed to create an estimated n(z) for each tomographic bin.

In Figure 2.6.2 we show the effect of stacking fiducial p(z)'s (orange curve), stacking broadened p(z)s (green curve) and the true underlying n(z) using spec-zs for three different tomographic bins. Differences in the means of the curves are evident by eye. Additionally, the overall shapes-and notably the widths-of the stacked fiducial p(z)s differ considerably from the true distributions. Although this PRIMUS sample is not necessarily representative of the entire DLS sample, the differences in the means and widths between the true n(z) and the stacked photo-z PDFs may lead to systematic errors further along in our analysis. The extent that the mean and width errors in tomographic n(z)s impact cosmology inference will be examined in the following section.

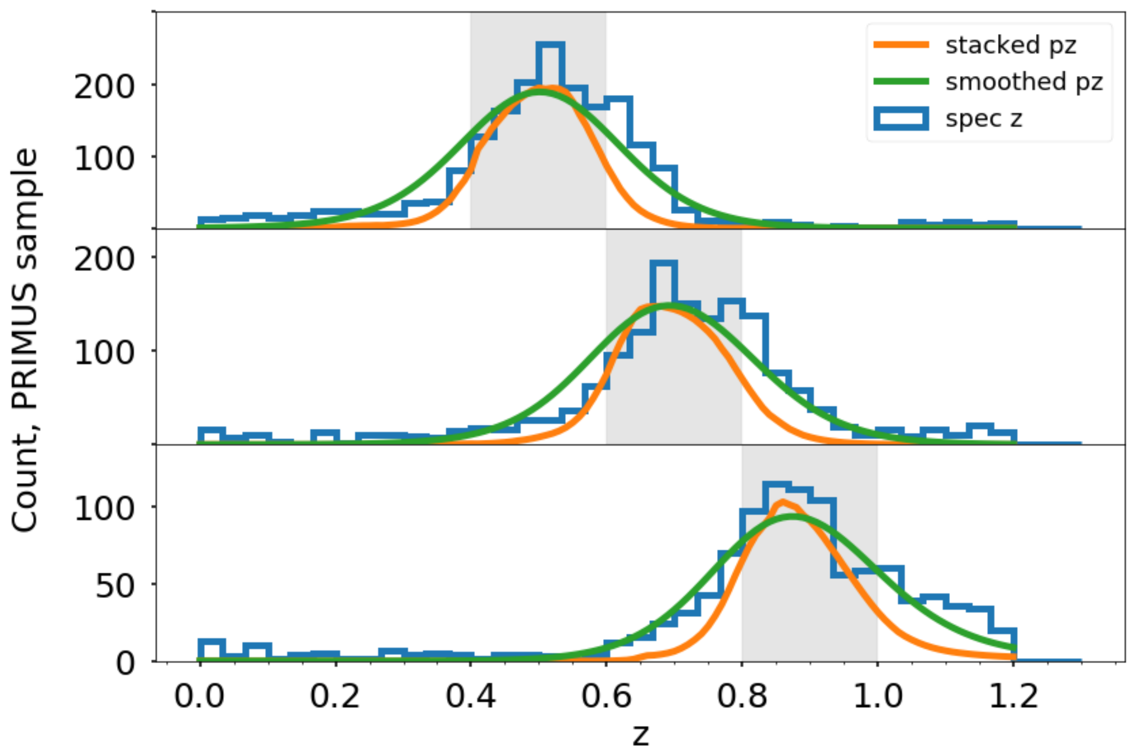

Figure 2.6.2: Different n(z) distributions for a sub sample of galaxies for which spec-z's are known. Galaxy selection and tomographic binning follows the procedure outlined in 2.3.3. Solid gray shaded regions indicate the redshift range that defines the tomographic bins. From top to bottom, the panels are for Source bins 0, 1, and 2. The orange curves show the n(z) from stacking p(z)s, as outlined in 2.3.3. The blue histograms show the corresponding n(z) using spec-zs. Discrepancies between the p(z) and spec-z n(z) mean and width are evident by eye. We attempt to mitigate the differences by convolving individual p(z)s with an optimal filter before stacking them, as discussed in 2.6.2, giving the solid green curves. The width correction techniques make the stacked p(z) more true to the true underlying n(z). The impact these n(z) errors have on cosmology inference outlined in 2.7

The fiducial p(z)s were broadened using a Gaussian kernel to create the corrected p(z)s. The choice of a Gaussian is perhaps somewhat naive in this instance. Examining the top two panels, one can see it is the *higher* redshift end of the fiducial p(z)s which need to be stretched, while the the lower redshift end of the distributions more or less agree with the true underlying n(z) and do not require correction. Stretching only one end of the fiducial p(z) cannot be achieved with a symmetric kernel, like the Gaussian. In principle, one can use the original PIT histogram to construct a transformation that enforces the new PIT histogram be uniform (Bordoloi, Lilly, & Amara, 2010). However, given the small sample size of the PRIMUS galaxies, we risk over-fitting if we use too complex a transformation. Additionally, even though our approach of using a Gaussian filter is not optimal, the new p(z)s are nonetheless an improvement over the fiducial p(z)s, as can be seen by comparing the two PIT histograms and as measured by the KL divergence.

2.7 Joint Analysis and Cosmology Inference

2.7.1 Co-variance Estimation

Cosmology inference requires a covariance matrix in addition to the data vector. Calculating the covariance for the cosmological probes is still an active area of research; see Krause et al. (2017) and references there in for a full summary. There are several families of techniques that have been used in the literature to obtain a robust covariance matrix in weak lensing studies. These include internal estimators, analytically calculated covariances, and numerical simulations. Internal estimators like the Jackknife and Bootstrap technique re-sample the observed data set in order to calculate covariances. This allows for effects like depth, field geometry, field masking, shape noise, and other survey specific systematics to be considered while calculating the covariance. Jackknifing has been found to be accurate provided the scales of the angles being probed by our two–point correlation functions are the size of or smaller than the jackknife regions. There are competing needs when using the jackknife: on one hand we have the need for the number of jackknife regions to exceed the number of data points, and on the other we would like to make the regions smaller to allow for more regions (Singh, Mandelbaum, Seljak, Slosar & Vazquez Gonzalez, 2017). Indeed, the small survey of the DLS makes jackknifing intractable. The small area does not permit a sufficient number of jackknife regions while simultaneously keeping the regions large enough to be bigger than the angular scales being considered by our two point probes.

Analytic covariances often involve the assumption of a Gaussian density field. This assumption is valid in the linear regime on large scales. Analytic covariances are computationally less expensive to calculate and do not contain statistical noise the other two strategies are susceptible to (Krause et al., 2017). However, because of the small, complex, footprint of the DLS and the small shape noise of the survey, a Gaussian covariance matrix may not be best suited to this survey (Chang et al., 2019).

Simulations can allow for some survey specific systematics-like field geometry, an approximate description of shape noise, exclusion regions due to bright stars etc to be considered. If many mock surveys are created, the covariance matrix can be obtained by calculating the covariance of the data vectors from each mock. A caveat to bear in mind is that super-survey modes will not be captured by these simulations. We use the publicly available package Full Lognormal Astro-fields Simulation Kit (FLASK) to simulate the DLS 549 times (Xavier et al., 2016). By providing input power spectra for all possible combinations of convergence and matter density fields, cosmological parameters, and angular and radial selection functions, FLASK can simulate the DLS at a catalog level. The catalogs include galaxy angular position on the sky, redshift, and galaxy shape after gravitational lensing. This allows us to measure cosmic shear, galaxy clustering, and galaxy galaxy lensing, using the same pipeline used for the real DLS catalogs. The covariance of these data vectors is calculated and used as the covariance matrix for our cosmological probes.

2.7.2 Likelihood Analysis

Ultimately we wish to obtain constraints on cosmological parameters that define a cosmological model. We will proceed by using a Bayesian framework, which will allow us to produce a posterior distribution for cosmological parameters of interest. Formally, the posterior is defined as

$$P(\mathbf{Y}(\theta)|\mathbf{X}) = \frac{P(\mathbf{X}|\mathbf{Y}(\theta))P(\mathbf{Y}(\theta))}{P(\mathbf{X})} \tag{2.7.1}$$

where $\mathbf{Y}$ is the a model with parameters θ, $\mathbf{X}$ is the observed data, $P(\mathbf{X}|\mathbf{Y})$ is the likelihood, $P(\theta)$ is the prior, and $P(\mathbf{Y}|\mathbf{X})$ is the posterior

In principle one can use this framework to calculate posteriors for parameters in different cosmo-

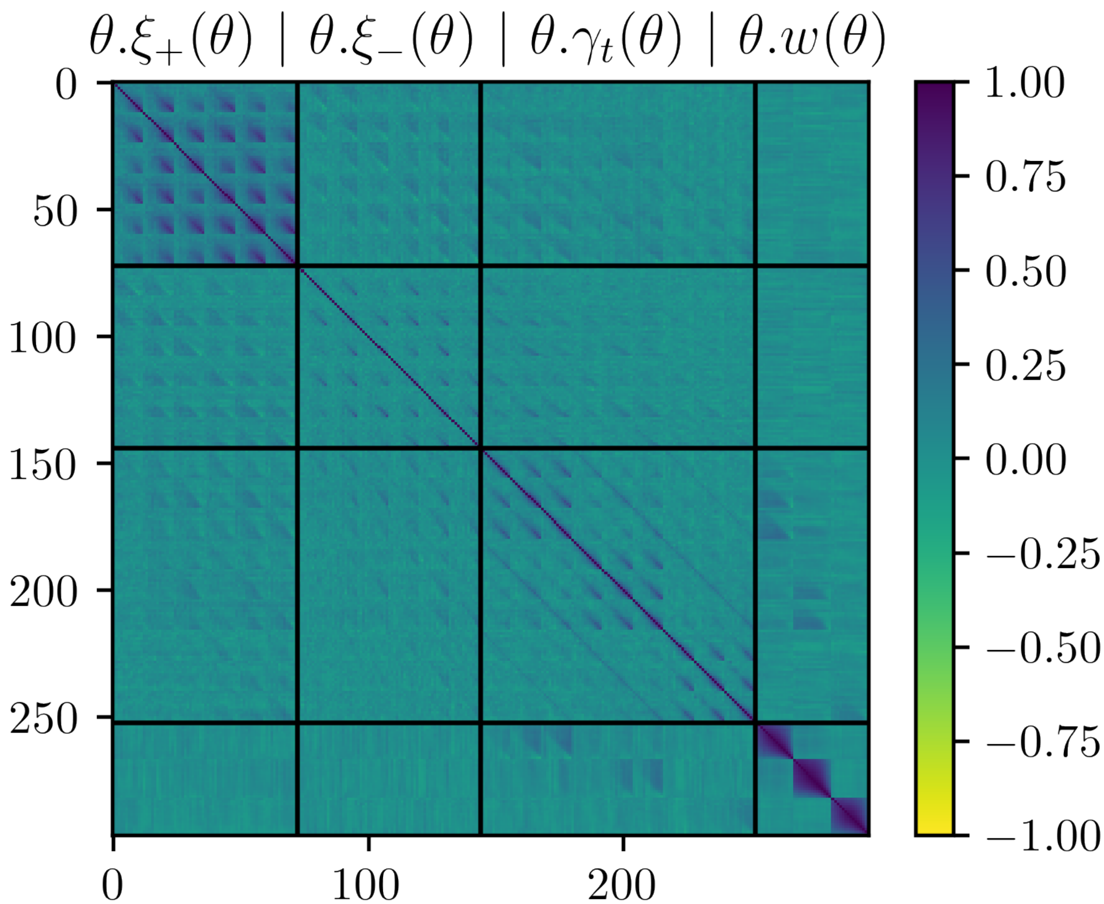

Figure 2.7.1: The full covariance matrix for three cosmological probes, cosmic shear, galaxy-galaxy lensing, and galaxy clustering. The covariance matrix is constructed by emulating the DLS in 549 FLASK simulations.

Table 2.2: Summary of nuisance, astrophysical, and cosmological parameters we marginalize over in our likelihood analysis. The photo-z mean parameters are determined by comparing photo-zs from subset of our data for which we we have spec-zs. photo-z biases have Gaussian priors, and remaining parameters have uniform priors. For Gaussian priors, the third column represents the mean and standard deviation of the prior. For uniform priors the column represents the edges of the flat tophat prior.

heightparameter	prior parameters
Nuisance Parameters	
L0 photo-z bias	0 0.02
L1 photo-z bias	0 0.02
L2 photo-z bias	0 0.02
S0 photo-z bias	0 0.02
S1 photo-z bias	0 0.02
S2 photo-z bias	0 0.02
Shear calibration	-0.03 0.03
Astrophysical Parameters	
I.A. Amplitude	-5.0 5.0
Galaxy bias	0.8 3.0
Cosmological Parameters	
Ω_m matter density	0.1 1.0
n_s spectral index	0.8 1.2
Ω_b baryon density	0.03 0.06
σ_8 power spectrum normalization	0.2 1.4
h Hubble parameter	0.5 .85

logical models. Indeed, Albrecht, et al. (2006) argue that a joint combination of several cosmological probes is a promising avenue to measuring dynamic dark energy and modified gravity. In this work, however, we will be examining a flat ΛCDM cosmological model.

Given a model cosmology with parameters θ the likelihood function is given as

$$L = \frac{1}{(2\pi)^{m/2}|C|^{1/2}}\exp[-.5(\mathbf{X}-\mathbf{Y}(\theta))^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{X}-\mathbf{Y}(\theta))] \tag{2.7.2}$$

where $\mathbf{X}$ is the data vector, $\mathbf{Y}(\theta)$ is the predicted data value assuming a flat ΛCDM cosmology, $\mathbf{C}$ is the covariance matrix and m is the dimension of the data vector. In practice, it is more practical to use the log likelihood to avoid computational problems, and this is true in our work as well. As a consequence, $P(\mathbf{X})$ will become an additive constant we may neglect. In fact, it can be seen in Equation 2.7.1 that $P(\mathbf{X}$ does not matter for model selection, weather we take the log or not.

The likelihood function definition shows that, in principle it is strictly speaking the *inverse* of the covariance matrix, the so called precision matrix, that is used, and not the covariance matrix itself. Our FLASK covariance matrix is an unbiased estimator of the true covariance matrix, but it none the less contains noise. Because of the noise contribution, upon inversion the estimated precision matrix itself is not an unbiaesd estimator of the true precision matrix. We can mitigate this effect, however, by applying the so called Hartlap Correction Factor, α_A where $\mathbf{C}^{-1} = \alpha_A\hat{\mathbf{C}}{-1}$ (Hartlap J., Simon P., Schneider P., 2007).

We use the publicly available code `COSMOSIS` to carry out a Monte Carlo Markov Chain to sample the posterior. At a high level, the predicted data vectors $\mathbf{Y}$ are generated by first assuming values of the model parameters θ, which are sampled from their priors. The θ are used to calculate the full three dimensional non-linear matter power spectrum using `CLASS` (Kilbinger, et al., 2009)and `HALOFIT` (Takahashi, Sato, Nishimichi, Taruya & Oguri, 2012). After, the Limber approximation is applied to the matter power spectrum to make a two dimensional matter power spectrum. The power spectrum of the different two point statistics (see eq 2.2.1, 2.2.4, and 2.2.3) are derived from $P\left(\frac{l+1/2}{f(x)}\right)$ and depend on the normalized n(z) of the tomographic bins being considered. A Hankel Transform is applied to the power spectra to generate prediction values in real space, which are arranged to form the predicted data vector $\mathbf{Y}(\theta)$.

2.7.3 Priors

We define several nuisance parameters which we marginalize over during our likelihood analysis. We use Gaussian priors for all n(z) priors, defined by their mean and standard deviation. The mean and standard deviation are set to 0 and 0.02 respectively, for these priors. The standard deviation are determined by comparing a sample of photometric redshifts from galaxies which also have spectroscopic redshifts from PRIMUS, where the typical scatter is found to be approximately 0.02.

The multiplicative galaxy bias, b_i for each lens bin is of particular importance to our study. In addition to its high degeneracy with Ω_m and σ_8 (Prat et al., 2018), we are examining any degeneracy it may have with n(z) shape errors. We use uniform priors from 0.8 to 3 for each galaxy bias. The large range in prior values avoids any prior bias in our likelihood analysis.

We use a flat prior in the range of -0.03 to 0.03 to marginalize over shear calibration errors with a uniform prior, following the method of Yoon et al. (2019) and Jee et al. (2016). The multiplicative shear bias for the ith source bin, m_i, is defined as $\gamma_t^{'} = (1 + m_i)\gamma_t$.

We also define several astrophysical parameters to marginalize over. Intrinsic alignments of galaxies can mimic a coherent alignment of galaxy shapes, contaminating the cosmic shear signal, and are a key systematic is cosmic shear studies. We adopt an intrinsic alignment model which has amplitude dependence, A_{IA}. Following Yoon et al. (2019), we begin with the model adopted in Catelan, Kamionkowski, & Blandford (2001)

$$P_{\delta I} = -A_{IA} C_1 \rho_c \frac{\Omega_m}{D(z)} P_{linear}(z) \tag{2.7.3}$$

where A_{IA} is the dimensionless amplitude $C_1 = 5 \times 10^{-14} h^{-2} M_\odot Mpc^3$, $P_{linear}(z)$ is the linear matter power spectrum, ρ_c is the critical density of the universe today, and $D(z)$ is the growth factor at z normalized to be one at $z = 0$. Following Bridle & King (2007) and Yoon et al. (2019) we exchange the linear matter power spectrum with the non-linear version. This is used to construct the power spectrum of the intrinsic alignment contribution to the galaxy-galaxy lensing power spectrum

$$P_{IA}^{ij} = b \int_0^{\chi_h} \mathrm{d}\chi \frac{p_g^i p_\kappa^j}{f^2(\chi)} P_{\delta I}\left(\frac{l + 1/2}{f(\chi)}; \chi\right) \tag{2.7.4}$$

We marginalize over A_{IA} with flat priors in the range -5 to 5.

Finally, we marginalize over cosmological parameters, all with uniform priors which encompass values obtained from recent studies. The ranges for these flat priors are $0.1 \leq \Omega_m \leq 1.0$; $0.5 \leq h \leq 0.85$; $0.2 \leq \sigma_8 \leq 1.4$; $0.8 \leq n_s \leq 1.2$; $0.02 \leq \Omega_b \leq 0.06$.

2.8 Results for Galaxy Clustering + Galaxy-Galaxy Lensing

2.8.1 Effect of Broadening N(z)

In Figure 2.8.1 we show the parameter constraints on the cosmological parameters Ω_m, σ_8 and the derived parameter $S_8 \equiv \sigma_8(\Omega_m/.3)^{.5}$ for a flat ΛCDM cosmology. The degeneracy between Ω_m and σ_8 motivates the definition of S_8, which splits this degeneracy, as can be seen in the Figure. S_8 is also useful for comparing the results from different studies.

The results for the uncorrected set up, where we use the original n(z) distribution before broadening with a trained filter, are shown in the blue contours where the darker region shows the 1-σ interval and the lighter shaded region shows the 2-σ region. The results for the corrected case, where we have broadened the n(z) in the method discussed in 2.6.2, are shown in cyan. For the corrected setup we constrain

$S_8 = 0.739^{+0.054}_{-0.050}$

$\Omega_m = 0.412^{+0.113}_{-0.094}$

$\sigma_8 = 0.601^{+0.088}_{-0.067}$

For the uncorrected result, we constrain $0.841^{+0.062}_{-.061}$. There is a $> 1\sigma$ tension between the two results. It warrants noting that the uncorrected case is in good agreement with the previous DLS results for cosmic shear $S_8 = 0.818^{+0.034}_{-0.026}$ (Jee et al., 2013) and the DLS results for power spectrum space galaxy clustering + galaxy-galaxy lensing $S_8 = 0.810^{+0.039}_{-0.031}$. The corrected result is in a $> 1\sigma$ tension with the previous DLS results.

2.8.2 Effect of Scale Cuts

In our analysis we made a choice on our scale cuts. The minimum scale is determined by model uncertainties in baryonic physics. The maximum scale is set by the footprint of the five fields in the DLS. In Jee et al. (2013) and Chang et al. (2019) the effects of how scale cuts on the DLS cosmic

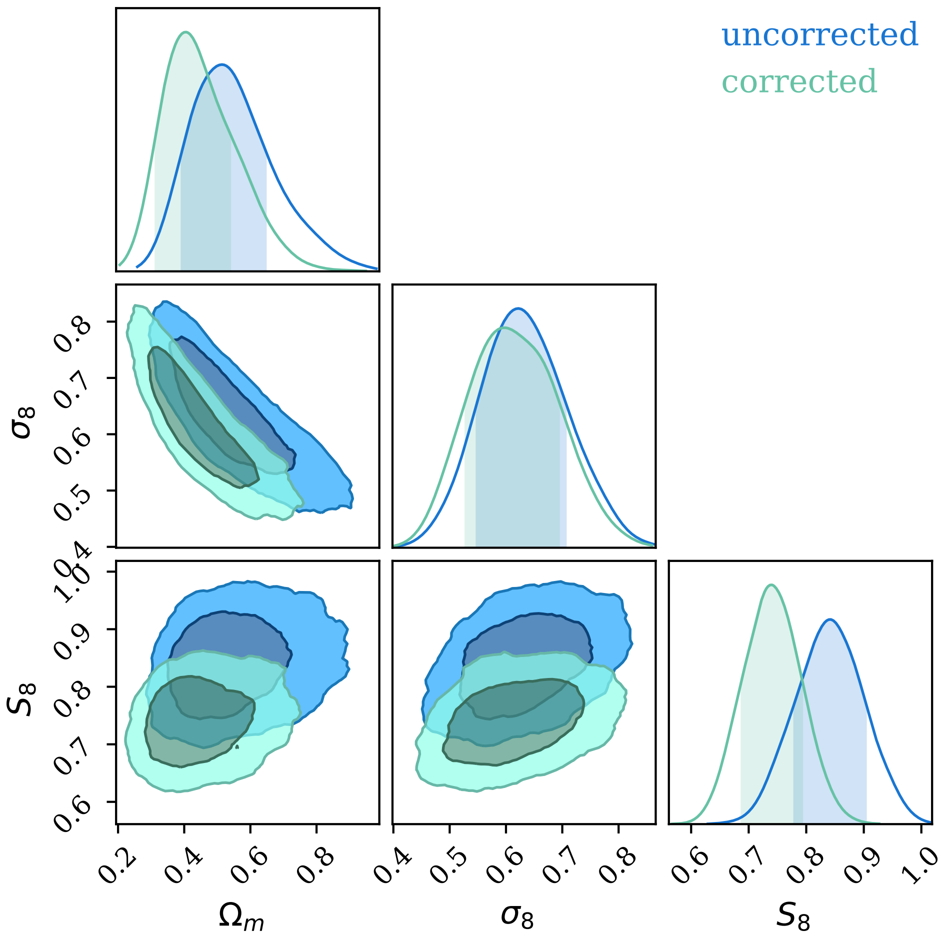

Figure 2.8.1: We present two cosmological constraints from combined galaxy clustering and galaxy-galaxy lensing probes. The uncorrected case, in blue, shows constraints where we have not attempted to correct the widths of the n(z) distributions in our samples. The corrected case, in cyan, shows constraints after we have corrected the widths of the n(z) distributions in our tomographic bins. The result of correcting the n(z) is for the confidence interval to 'slide down' in the $S_8 - \Omega_m$ plane, ultimately shifting the constraint on the growth of structure parameter S_8 to lower values.

shear data severely impacted the cosmological constraints are presented. For example, when using conservative scale cuts where the minimum scale corresponds to 1.3 comoving Mpc in each redshift bin, the DLS constraint moves to a large values of Ω_m and S_8 to greater than .8 and .9 respectively.

With an eye towards this work, we investigate the effect of scale cuts on our work presented here. We compare three cases where we have unified scales to all probes: the first case where the minimum scale is 2 arcminutes, an intermediate case where the minimum scale cut is 6 arcminutes, and a conservative case where the minimum cut is 20 arcminutes. All maximum scale cuts are set to 90 arcminutes. All data vectors make use of the corrected n(z).

In Figure 2.8.2 we present the constraints of the different scale cut cases. The corrected case (2 - 90 arcminutes), intermediate case (6 - 90 arcminutes), and conservative case (20 - 90 arcminutes) are shown in cyan, orange, and purple/blue respectively. Several comments are in order. The conservative case noticeably loses constraining power relative to the other two cases. This is to be expected given the dramatic reduction in data that is being used in cosmology inference. Enlarged contours in the $\Omega_m - S_8$ planes are also presented in (Chang et al., 2019) when conservative scale cuts are used on other surveys. However, in addition to broadening, the conservative case moves the contour up to higher values of S_8, exceeding 1. The intermediate case, on the other hand, seems to follow the direction of the 'banana' in the $\Omega_m - S_8$ plane, sliding down to high values of Ω_m and small values of σ_8.

While the corrected case and intermediate case ultimately constrain S_8 values that are consistent with one another, we must indicate that the constrained values of Ω_m and σ_8 differ at the 1σ level. Furthermore, the intermediate case gives values that are drastically different than those presented in the recent literature.

As noted above, the conservative case presents as an outlier in the parameter S_8, constraining values between approximately 1 and 1.25. This value is in significant tensions with the other cases, and other values presented in the modern literature.

Ultimately, the key take away from Figure 2.8.2 is that different scale cuts lead to different cosmological constraints. This is a troubling observation, as it indicates some as of yet uncharacterized systematics still remain in the data. For a truly robust data set, the anticipated outcome of progressively moving the minimum scale to higher angles would be looser-but *consistent* constraints.

To illustrate the point that scale cuts should lead to consistent but looser cosmological con-

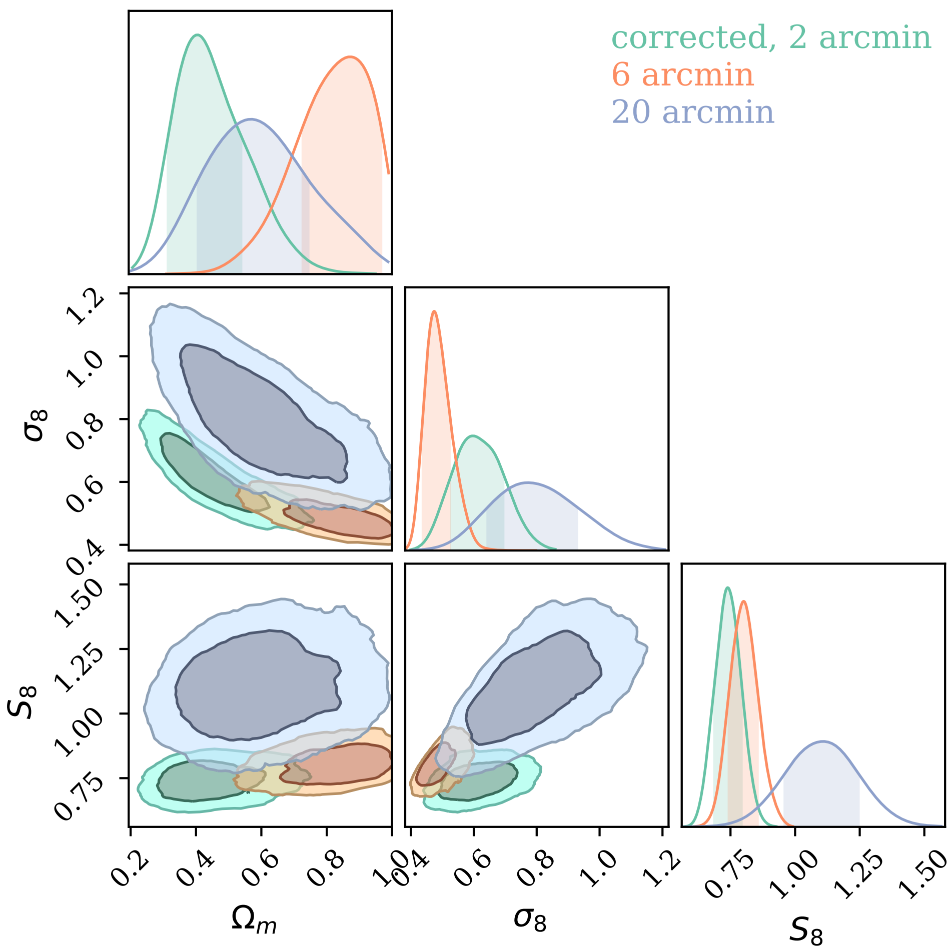

Figure 2.8.2: We present 3 cosmological constraints, each using different scale cuts in our galaxy clustering and galaxy-galaxy lensing data vector. The scale cuts are uniform across probes. Specifically, the corrected case (2-90 arcminutes), intermediate case (6-90 arcminutes) and conservative case (20-90 arcminutes) are shown in cyan, orange, and purple/blue respectively. The constraints are highly sensitive to scale cuts. Namely, different scale cuts on the same data vector produce different cosmological constraints. This likely indicates unresolved systematic effects in the data remain.

straints, we turn our attention to The Dark Energy Survey (DES) (Troxel et al., 2018). The DES has provided their data vector and covariance matrix for their 3x2 point study, along with their cosmology inference pipeline. First, we demonstrate that our "g3l" inference pipeline and the published results are consistent with one another. We configure our pipeline to match the same scale cuts and priors as those described in Abbott et al. (2018). In Figure 2.8.3 we show the results of our pipeline (green) and the published DES result (blue) for galaxy clustering and galaxy-galaxy lensing. We find the two are in good agreement, as the contours in the three different planes sit precisely atop one another, and the marginalized posteriors.

Now that we have demonstrated that we can recover the published results, we examine the effect of making scale cuts to the data on the resulting cosmology. We run our g3l pipe inference, again using the DES data vector, covariance matrix, and preserving the same priors as before. However this time we introduce scale cuts to the data. We take the most conservative scale cuts that were used in all galaxy-galaxy lensing bins-64 arcminutes to 250 arcminutes-and apply them across the sample. Likewise, for galaxy clustering, we apply the most conservative scale cut used across the whole sample, this time 43 arcminutes to 250 arcminutes. The impact this has on cosmology inference is show in the red contours and curves in Figure 2.8.3. We can see the constraints on Ω_m are largely unchanged, as there is superb agreement between all three marginalized posteriors. Examining the contours, we can see the 1 sigma confidence interval is broader in the red contour than the blue and green. This trend holds when looking at the marginalized posteriors for σ_8 and S_8 where we see the 1 sigma regions of the red curves contain and go slightly beyond the 1 sigma regions of the blue and green curves.

2.9 Effect of Large Angular Scales

In Figures 2.4.2 and 2.4.1 we can see a high deviation between the amplitude of the mesured data at high angular scales and the best fit model cosmology. To examine their contribution to the cosmology results, we take the corrected data vector and remove the data points at the highest angular scale, and rerun our cosmology inference. The results are shown in Figure 2.9.1 in the green contour, while the original corrected data vector is shown in blue again. The results are highly consistent, suggesting they do not have a high impact on the final cosmology result.

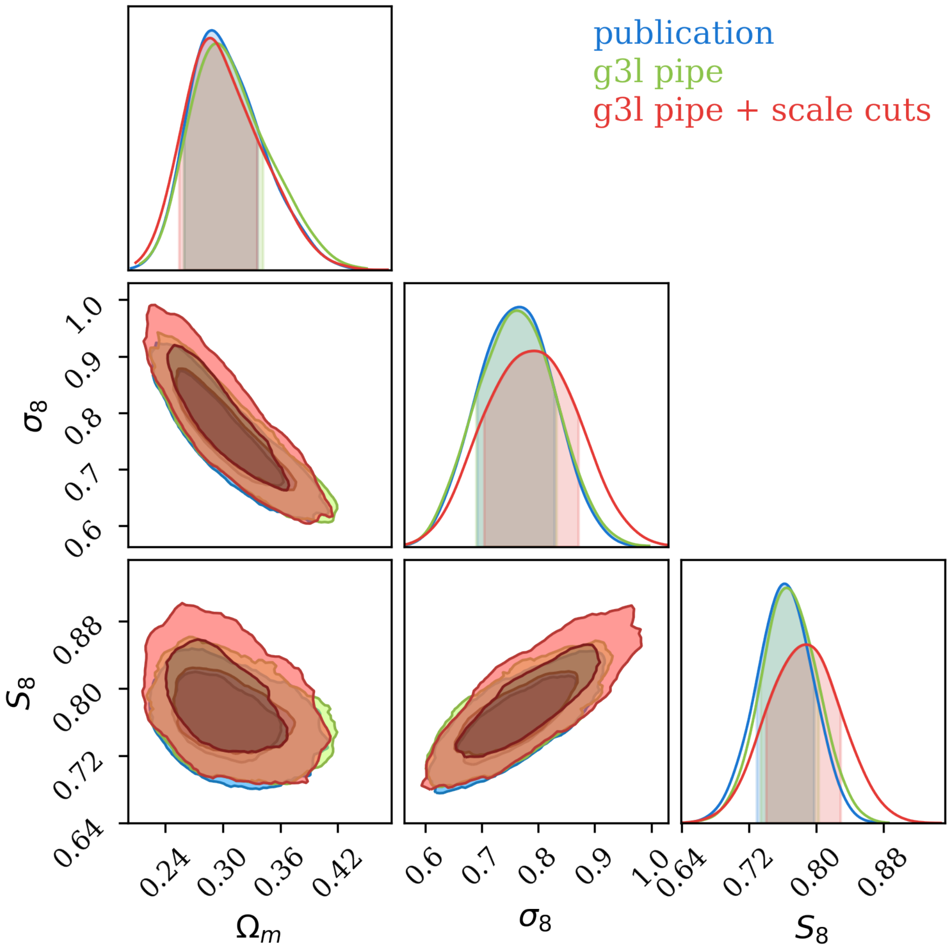

Figure 2.8.3: Cosmological constraints using Dark Energy Survey (DES) data vector and covariance. We show three different cases. First, in blue we show the published result from Abbott et al. (2018). Second, we demonstrate we can recover the published results using our pipeline by producing the green contours, which are in near perfect agreement with the published results. Third, we apply conservative scale cuts to the DES data vector. The scales used for galaxy-galaxy lensing are 64 - 250 arcminutes, and the scales used for galaxy clustering are 43 - 250 arcminutes. The resulting constraints are shown in red. Employing scale cuts produces contours and marginalized posteriors are consistent with, but less constraining than the published cases, which were done with more liberal scale cuts.

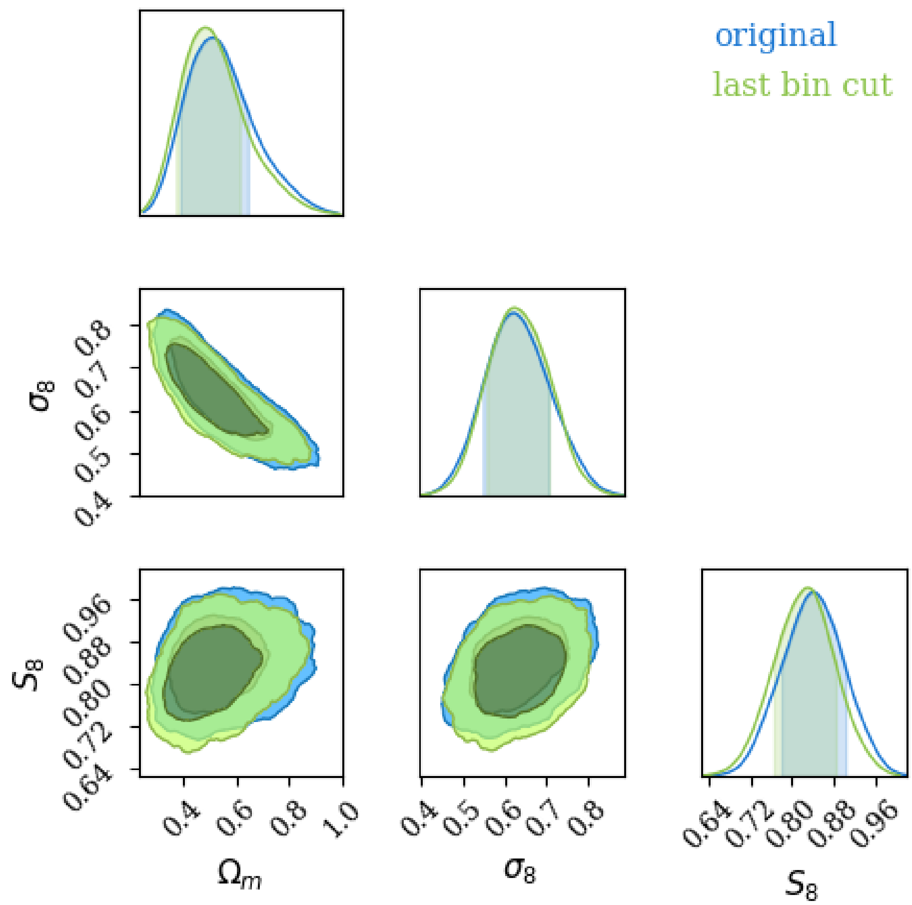

Figure 2.9.1: In blue the cosmology results from the corrected data vector, again shown in blue. In green, we show the cosmology results of removing the highest angular scale data points from all the correlation functions. The results are highly consistent.

2.10 Results For Cosmic Shear

In Figure 2.10.1 we show parameter constraints on the cosmological parameters Ω_m, σ_8 and the derived parameter S_8. The color schemes for the contours are unchanged with respect to 2.8.2 above, namely the corrected case (2 - 90 arcminutes), intermediate case (6 - 90 arcminutes), and conservative case (20 -90 arcminutes) are shown in cyan, orange, and purple/blue respectively. For the corrected setup we constrain

$S_8 = 0.756^{+.040}_{-.039}$

$\Omega_m = 0.163^{+.036}_{-.034}$

$\sigma_8 = 1.002^{+.103}_{-.093}$

In Figure 2.10.1 we can see the effect that scale cuts have on our cosmology constraints. While the corrected case and intermediate case are consistent at the 2σ level, the conservative case's contours are far removed in all panels. We can see a general trend that more and more conservative scale cuts leads to placing contours at higher and higher values in Ω_m and consequently $\S_8$. While this trend was noted in Jee et al. (2016), the authors did not push up to a minimal scale cut of 20 arcminutes, only going up to 4 arcminutes. The effect of scale cuts was only considered to be minor since the impact on S_8 was consistent at the 1σ level. However, in our work we can clearly see the impact that higher scale cuts impose on cosmological constraints is well beyond the 1σ level.

2.11 Results For Combining all Probes

In Figure 2.11.1 we present the constraints from combining all three cosmological probes together (blue/purple contour). For reference we also show the previous results from above for galaxy-galaxy lensing + galaxy clustering (orange) and cosmic shear (cyan). As expected, the constraining power of combining all three probes leads to tighter constraints than using the probes separately. While the galaxy clustering + galaxy galaxy lensing constraint is consistent with the full 3x2 point constraint, there is a more than 2 σ tension between the cosmic shear result and both the full 3x2 point constraint and galaxy clustering + galaxy-galaxy lensing. This is unexpected, as a truly robust result would see consistency across the board, as we have seen in DES (Abbott et al., 2018) and KiDS (Heymans et al., 2020).

It is also notable that different constraints on the galaxy bias were achieved by the full 3x2

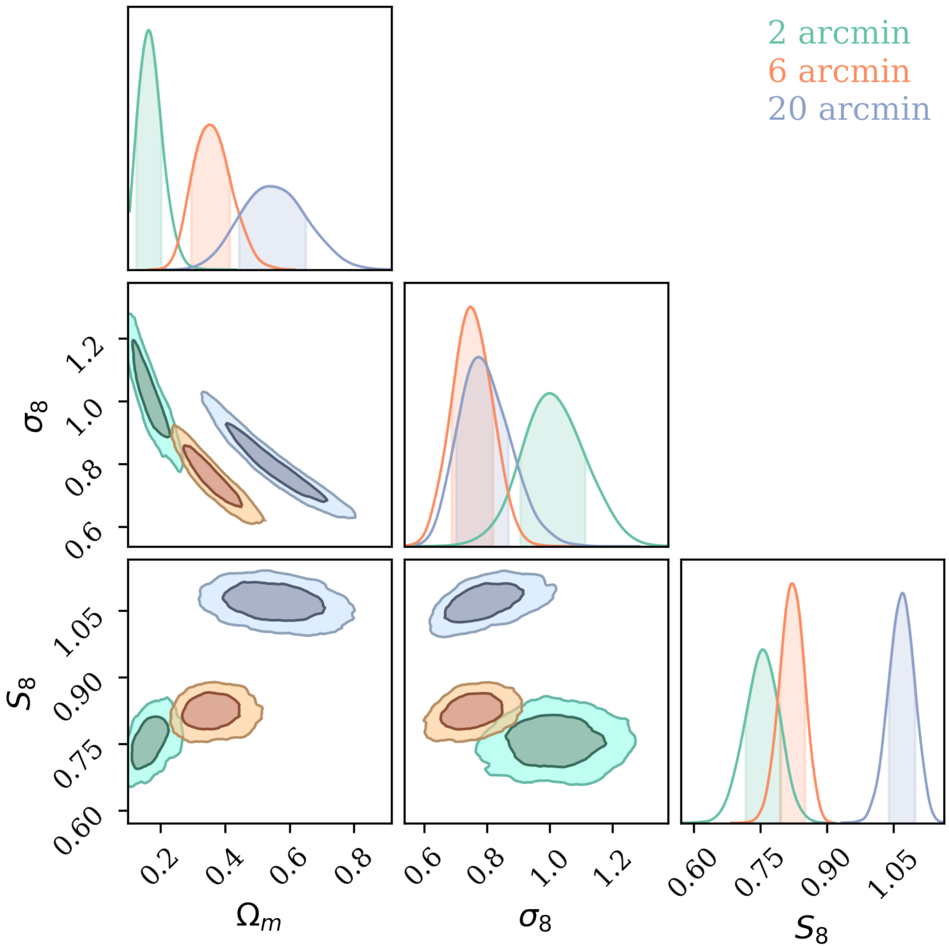

Figure 2.10.1: We present 3 cosmological constraints, each using different scale cuts in our cosmic shear data vector. Specifically, the corrected case (2-90 arcminutes), an intermediate case (6-90 arcminutes) and a conservative case (20-90 arcminutes) shown in cyan, orange, and purple/blue respectively. The results show sensitivity to scale cuts, where higher scale cuts appear to move the contours to higher values of Ω_m and consequently S_8. This likely indicates unresolved systematic effects in the data remain.

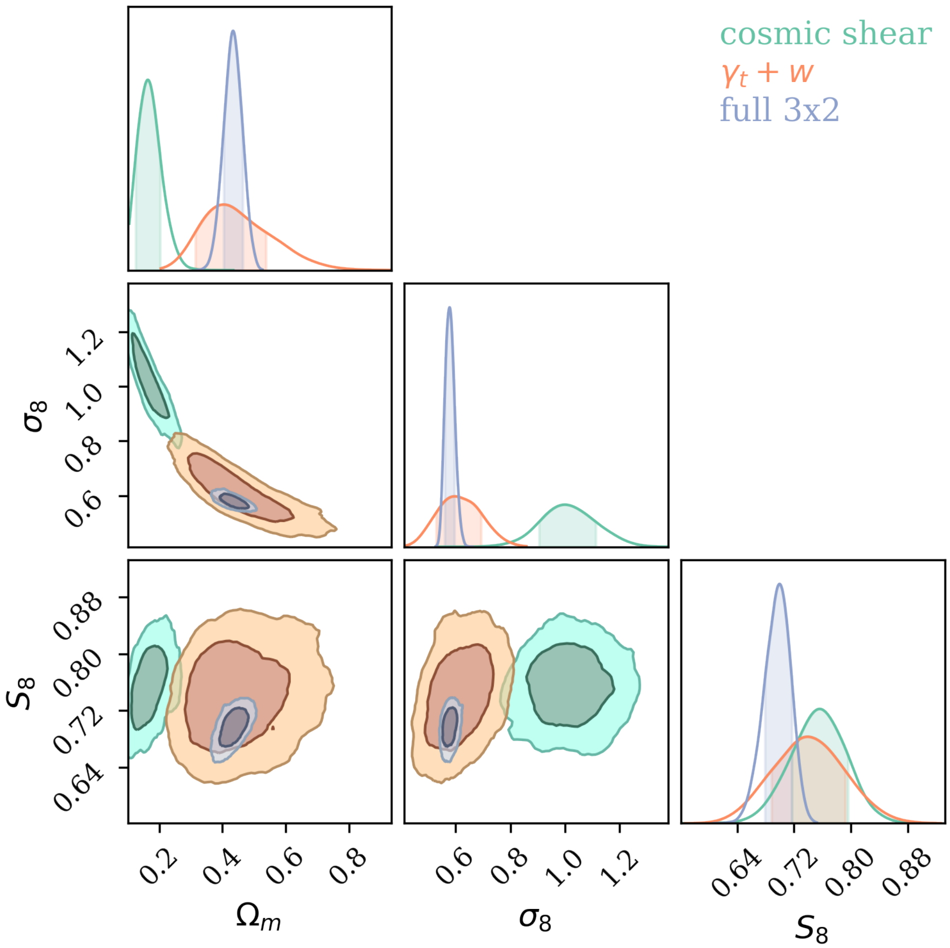

Figure 2.11.1: We present our cosmological constraints from cosmic shear, galaxy clustering + galaxy-galaxy lensing, and the result of combining all three two point functions (full 3x2pt) in cyan, orange, and purple/blue respectively. We can see the results from the full 3x2 point and galaxy clustering + galaxy-galaxy lensing are consistent with one another at the one sigma level. While the combined constraint is tighter than either of the separate two constraints-as expected- there is a 2 sigma tension between cosmic shear and both galaxy clustering + galaxy-galaxy lensing and the full 3x2 point constraint.

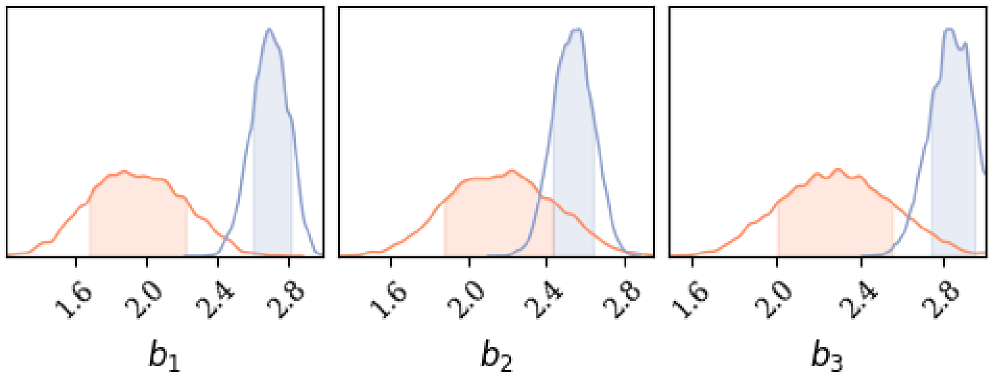

Figure 2.11.2: Marginalized constraints for the linear galaxy biases for three tomographic lens bins for galaxy clustering + galaxy galaxy lensing shown in orange, and for the full 3x2 point shown in purple/blue. The constraints are inconsistent with one another at the 1σ level.

point constraint and the galaxy clustering + galaxy-galaxy lensing. In Figure 2.11.2 we show the marginalized constraints on the linear galaxy biases, b_i for the three tomographic lens bins. The constraints from the galaxy clustering + galaxy-galaxy lensing are again shown in orange, and the constraints from the full 3x2 point constraint are shown in purple/blue. We can see the constraints between the two are inconsistent at the 1σ level. Referring back to Figure 2.11.1, even though the constraints on cosmological parameters are consistent with one another, this is not true of all fitted parameters. Taken all together, this implies that including the cosmological information from cosmic shear prefers larger galaxy biases.

To reiterate points made above, this points to some as of yet unidentified systematic sources of error in the data that remain. The results for cosmic shear and galaxy clustering + galaxy-galaxy lensing should be consistent with one another, and the joint combination of all three two point functions should be consistent with both of them if the data are robust.

2.12 Comparison of DLS S_8 with Other Studies

The surveys we compare to are DES Year 1 (Abbott et al., 2018), KiDS 1000 (Heymans et al., 2020), HSC Year 1 (Hamana et al., 2020; Hikage et al., 2018), *Planck* (Planck Collaboration et al., 2020), as well as previous studies from the DLS. In Figure 2.12.1, we show the resulting S_8 constraints from these studies. The first thing that draws the eye is the relative lack of constraining power in this work relative to the other studies presented. This is somewhat expected, as our especially conservative treatment of photo-zs lead to us discarding many galaxies in our sample. Our study relies on approximately a third of one million galaxies. Other DLS studies use approximately a million galaxies, and HSC, DES, and KiDS use several millions of galaxies, by comparison.

Overall we find good agreement between our study's cosmic shear, and galaxy clustering + galaxy-galaxy lensing probes and other weak lensing studies. The lack of constraining power in our results, however, means that we cannot weigh in on the so called 'S_8 tension', the tension in results for S_8 between some low redshift weak lensing surveys and *Planck*. For example, in Figure 2.12.1, the KiDS-1000 3x2pt result is in significant tension with the *Planck* result, while the previous DLS results are in good agreement with *Planck*. The results presented here, on the other hand, have large enough uncertainty that they are in agreement with weak lensing studies and under 1σ difference with *Planck*.

The full combined full 3x2 point has tighter constraints on S_8, as we expect due to the complimentary nature of the interlocking probes. While the full 3x2 point result is well within agreement with the results of our cosmology probes shown here, it is nonetheless in 1σ tension with other weak lensing results.

2.13 Possible Sources Of Systematic Error

We saw in previous sections we saw some evidence that some systematic errors may be unmitigated in our data. To try to tease these out, we can perform consistency tests and null tests-where there are expected result-as a way to stress test the integrity of our data set. For example, the shear ratio test measures the relative correlation between the shapes of galaxies and stars. This is expected to produce zero signal, since stars in the Milkyway halo are not lensed by the large scale structure of the universe, while higher redshift galaxies are (Jee et al., 2013). In Figure 2.13.1 we show the results of the shear ratio test for our sample. In the blue (orange) points we show the real (imaginary) part of the signal. Additionally, the expected signal amplitude, zero, is shown in a dashed black line. The y-axis has been scaled by angle so the amplitudes here can be directly compared to our cosmic shear and galaxy-galaxy lensing results. We find the real part of the signal is consistent with zero, while the imaginary part of the signal has non-zero amplitude at angular scales of about 4 and 80 arcminutes. The presence of a non-zero signal is perhaps a hint that the

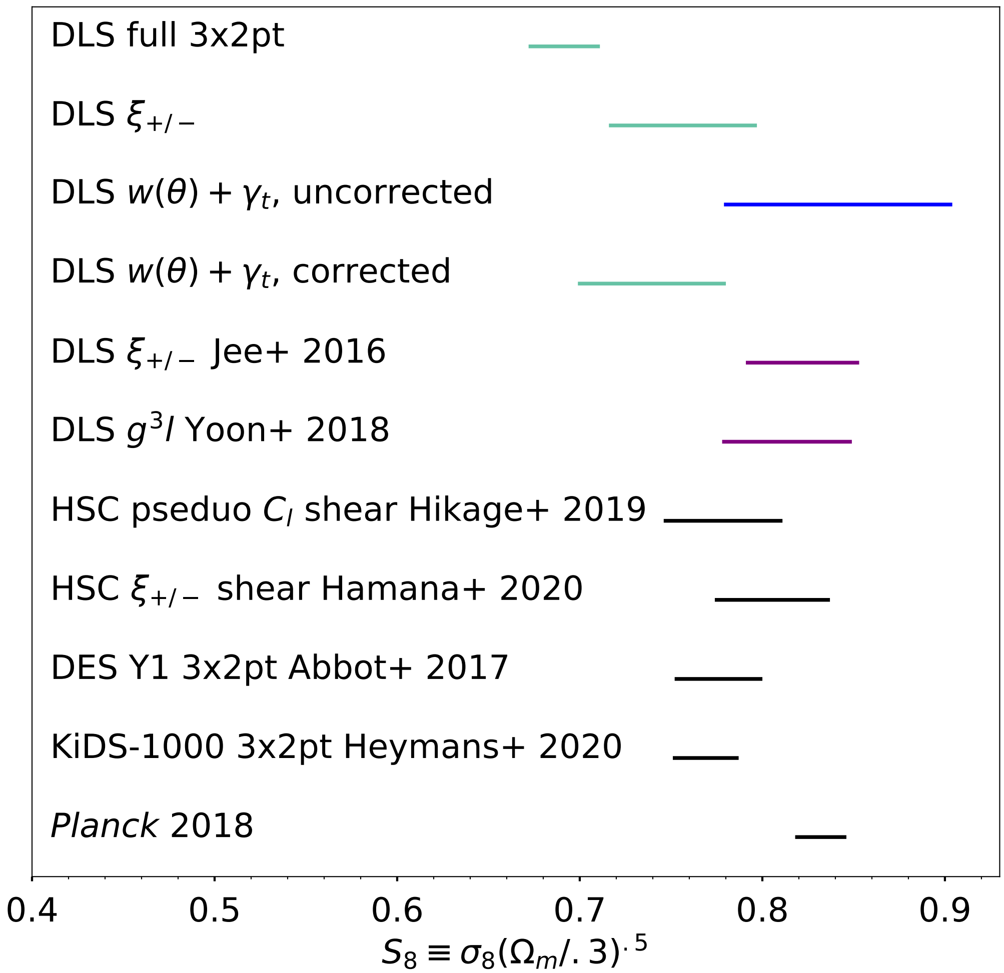

Figure 2.12.1: We show marginalized constraints on the growth of structure parameter $S_8 \equiv \sigma_8(\Omega_m/.3)^{.5}$ for our study, previous DLS cosmology studies, and recent constraints in the literature. The results from our study are shown in cyan and blue. The cyan results make use of the corrected n(z) while the blue result has the original uncorrected n(z). While they are not as constraining as other studies, our cosmic shear results and galaxy clustering + galaxy-galaxy lensing probes nonetheless find agreement within 1σ of other weak lensing results and *Planck*. The full 3x2point constraint, on the other hand, is constraining enough to show 1σ tension with other weak lensing studies, and a more than 2σ tension with *Planck*.

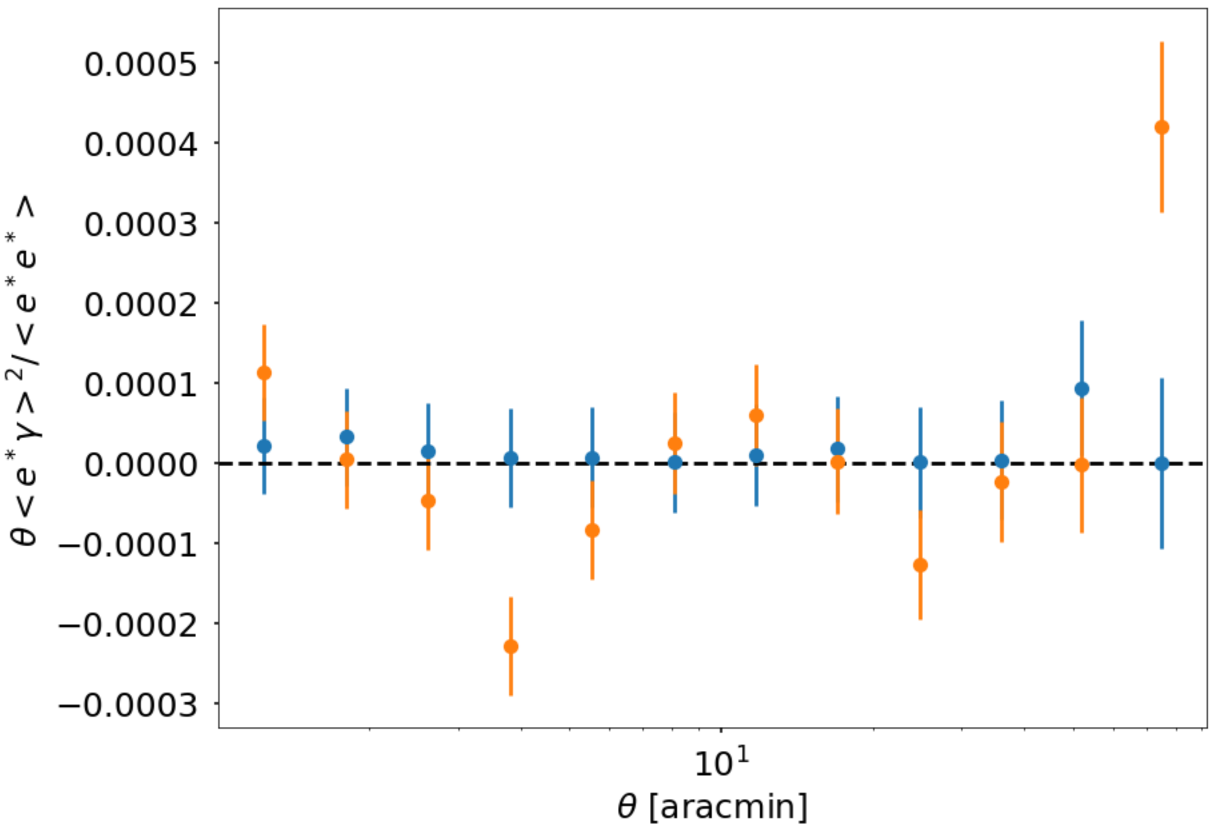

Figure 2.13.1: We show the shear ratio test, where we calculate the shape correlation between stars and galaxies and normalize by the auto-correlation of star shapes. This is a useful diagnostic of the additive shear bias, which has been taken to be zero in our analysis. In blue points we show the real part of the correlation, while in orange we show the imaginary part of the correlation. We have also multiplied the signal by its corresponding angular scale, so that the signal here can be compared to the signal in our shear-shear correlation functions. The real part is consistent with zero across all angular scales. The imaginary part is largely consistent with zero, except for scales around 4 and 80 arcminutes, suggesting perhaps the additive shear constant was non-zero for our sample.

additive shear constant should not have been taken to be zero for the sample.

Our null test indicates there are systematics involving the shape measurement on select angular scales. We may perform another test to isolate out any systematics associated with shape measurement by examining the angular two point correlation function, $w(\theta)$. Namely, if we fix the galaxy bias to some fiducial values and perform our cosmology inference with different angular cuts on the $w(\theta)$ data vector, we can separate out any dependence on angular scales and shear measurement, since $w(\theta)$ does not depend on galaxy shear, only galaxy position. While in practice one must marginalize over the galaxy bias, treating it as a nuisance parameter when doing cosmology inference, we must fix the galaxy bias in order for our results to have any constraining power for the purposes of our test. That is to say, the test may be elucidating our understanding of different systematics as a function of angular scale, but won't give us a bona fide cosmology constraint, as the galaxy bias is highly degenerate with Ω_m.

In Figure 2.13.2 we show marginalized constraints on cosmological parameters where we have performed such a test. There are three results, the first contour shown in cyan shows results when using the entire $w(\theta)$ data vector, where the smallest scale is 2 arcminutes. In orange we show the result of using 6 arcminutes as the smallest angular scale, and in blue/purple we show the results of using 20 arcminutes as the smallest angular scale. The results are all consistent with each other at the 1σ level, the only difference being a loss of constraining power as the inner angular scale is cut to higher and higher values. This is to be expected, as we are removing data from our analysis with our angular cuts.

The consistency across angular scale cuts is in contrast with Figure 2.10.1 and 2.8.2, where the constraints for our cosmic shear and galaxy clustering + galaxy-galaxy lensing move around the different planes depending on the scale cuts used. Since the $w(\theta)$ test is in isolation of any shear measurement systematics, this suggests the source sample being used in this work is infected with a particular systematic that affects shear correlations on different angular scales.

2.14 Conclusion

We utilize a Rubin Observatory precursor, the Deep Lens Survey, to investigate methods to mitigate errors on the n(z) shape can impart on cosmology constraints when using the combination of cosmic shear, galaxy clustering and galaxy-galaxy lensing probes. To correct for the width of the n(z), we use a validation set of data for which we have spectroscopic data, and empirically broaden individual p(z)s with a best-fit Gaussian filter. The width of the Gaussian filter is selected by minimizing the Kullback-Leibler divergence between the probability integral transform of the photo-z PDFs and the uniform distribution. We find that correcting the n(z) in this manner results in moving the constraint on the growth of structure parameter S_8 to lower values. The cosmic shear result is not affected by the broadening of the n(z), because of the broadness of the lensing kernel. By themselves, the cosmic shear result and the galaxy clustering + galaxy-galaxy lensing results are

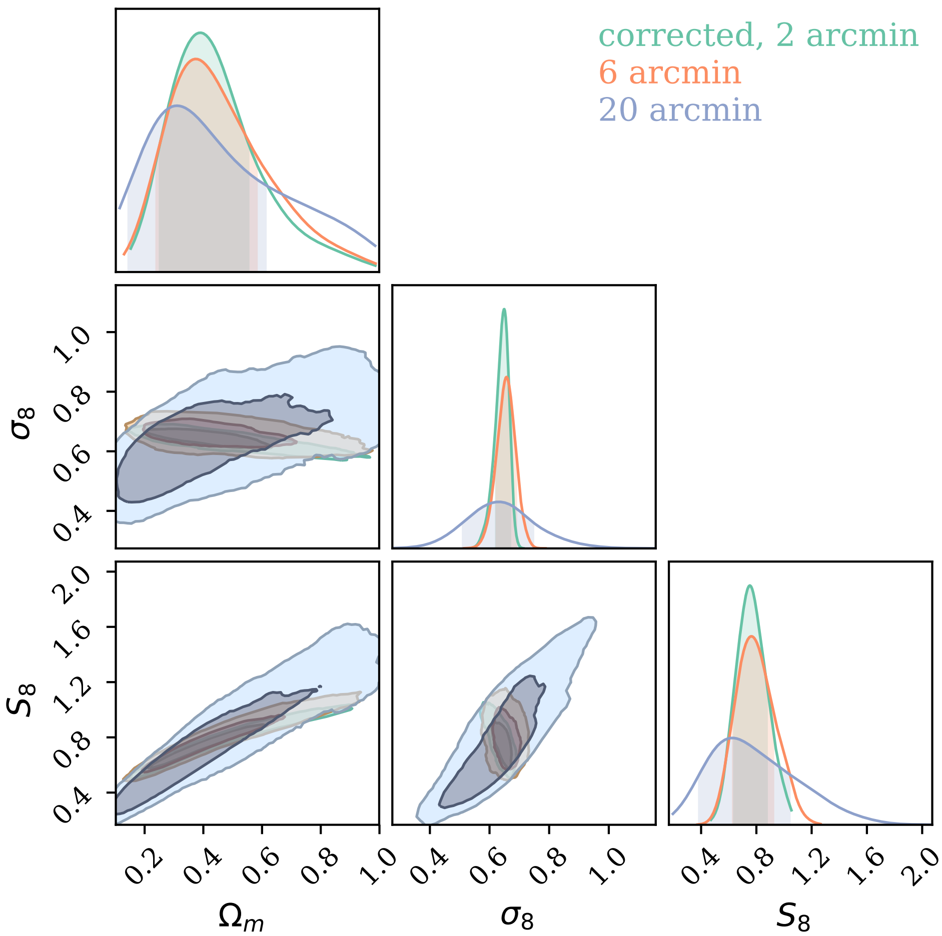

Figure 2.13.2: We show three different constraints on cosmological parameters for $w(\theta)$ where we have fixed the galaxy bias. While this will not deliver a robust measurement of cosmological parameters because we have fixed the galaxy bias and they are highly degenerate with Ω_m, it is nonetheless a useful diagnostic test. The three constraints are for our original corrected data vector of $w(\theta)$ (cyan), the data vector after a scale cut of 6 arcminutes, (orange), and the data vector with a scale cut of 20 arcminutes (blue/purple). While the constraining power is reduced the more restrictive the scale cuts are, the constraints are in remarkable agreement with each other. This is in contrast with Figure 2.10.1 and Figure 2.8.2, where the constraints are a strong function of angular cuts used. The results shown in this Figure are in isolation of any shear measurement systematics, and the robustness of the result here indicates there is likely a systematic associated with shear measurement that is a function of angular scale.

not as constraining as other weak lensing measurements, such as KiDS and DES, of S_8. None the less, we find agreement with them, and are less than 1 σ away from *Planck* constraints. We gain considerable constraining power by combining cosmic shear, galaxy clustering and galaxy-galaxy lensing, colloquially referred to the 3x2 point combination. Although the 3x2 point combination is in agreement with the cosmic shear alone and galaxy clustering + galaxy-galaxy lensing alone results, it is in significant tension with past DLS constraints, other modern weak lensing results, and a more than 2σ tension with *Planck*. We investigate possible sources of systematic error by performing null tests, and speculate there are unresolved systematics in the DLS that print through to different angular scales related to shape measurement.

2.15 Future Work

While our data set presents us with realistic photometric errors and photo-z errors, they are also subject to many systematic effects, some of which may contribute to the n(z) galaxy bias degeneracy. Truly understanding the impact these degeneracies impose on cosmology inference requires they are examined in isolation of all other sources of error. This is perhaps intractable in experimental data and requires a realistic simulation to fully evaluate. This will be the focus of our next paper, where we will repeat the analysis presented in this paper on the DESC Data Challenge 2 truth galactic catalog. Simulated data will also allow us to examine techniques to mitigate the effects of photo-z width and skew in isolation of all other effects. It may be possible to include photo-z width and skew as nuisance parameters which can be marginalized over, for example.

Chapter 3

The Effect Of Blending On Photometric Redshifts

3.1 Introduction

Weak lensing measurements for cosmology studies often require carefully detecting galaxies, measuring the shear they undergo, and calculating their distances. Detecting galaxies in the night sky is a familiar problem, and many automated algorithms have been developed to detect galaxies. Typically, detection of galaxies and other astrophysical sources involves passing a detection kernel over the image of interest making a smoothed image. The detection algorithm then identifies contiguous regions which are above a user defined detection threshold (Bertin & Arnouts, 1996; Bosch et al., 2018; Jarvis & Tyson, 1979). However, deeper weak lensing surveys now suffer more from galaxy blending the further down the detection limit they push, because of the increased number density of sources on the sky (Dawson et al., 2016).

Deblending is additionally a familiar problem to the community, and the same software developed for detection additionally preforms deblending. This is the process by which the light from different sources which overlap are disentangled. Often deblender algorithms look for different peaks which occupy the same above threshold regions in the smoothed images, and apportion flux in the region to the different peaks. However, in the case where the different peaks overlap sufficiently such that they cannot be individually identified, the deblender will not be called (Bosch et al., 2018). Thus, we can separate out blending into two regimes. One where a detection algorithm identifies multiple peaks in an above threshold region and a deblender is called to apportion the flux of the region to different sources, and another where the detection algorithm is unable to detect multiple peaks, and the deblender is not called. The former is referred to as a 'recognized' blend while the latter is referred to as 'ambiguous' or 'unrecognized blend' (Dawson et al., 2016).

Galaxy blending, in both the recognized and unrecognized regimes, may impose systematic errors on computing distances and shapes to blended galaxies. The most accurate way to determine the distance to a galaxy is perhaps with a spectroscopic redshift. However, selected samples of galaxies in weak lensing surveys are dominated by many millions of faint galaxies, for which obtaining a spectra would be prohibitively costly. As an alternative, multiband photometry is used to obtain a very coarse Spectral Energy Distribution (SED), which in turn is used as input for a statistical method to estimate a galaxy's redshift (Hildebrandt et al., 2010). However, the process of blending combines the flux of two distinct sources together, potentially introducing error to estimates of redshift if flux is not apportioned appropriately for recognized blends, and almost certainly introducing error to redshift estimation for unrecognized blends (MacCrann et al., 2020). To understand precisely how blending can impact photo-zs, we must first take a closer look at the methods to determining photo-z.

Photo-z methods broadly fall into two classes: template based approaches and empirical based approaches (Schmidt et al., 2020). Template based approaches at heart rely on computing a probability, p, between the predicted photometry, and the observed photometry. The predicted photometry comes from taking a library of sample SEDs, redshifting and scaling them, and convolving them with an observatory's throughput filter curves to obtain model predicted photometry. The negative log of the probability is equivalent to the χ square, and is computed as

$$\chi^2(z,T,a) = \sum_{i}^{N_{filt}} \left(\frac{f^i_{obs} - A f^i_{pred}(T,z)}{\sigma^i_{obs}} \right)^2$$

Here A is a normalization factor, $f^i_{pred}(T,z)$ is the predicted flux in the ith filter given a template T at redshift z, f^i_{obs} is the observed flux in the ith filter, and σ^i_{obs} is the observed flux error, and N_{filt} is the number of photometric filters.

By marginalizing over redshifts and templates, a probability density function (PDF) can be constructed, which encodes the likelihood of the galaxy being at a given redshift. In the blending regime, it is possible, for example, that photometry of two blended templates can be degenerate with another all together different template at a different redshift than either of the blended galaxies being considered.

Empirical codes for photometric redshift estimation vary in their algorithmic approach. A unifying theme is that these codes use magnitude, magnitude errors, and photometric colors as input, of 'features' in the machine learning parlance. Typically one uses a portion of data set aside to train the machine learning model, the training set. The training set must contain truth information, in this case spectroscopic redshifts are used, or other photo-zs with many photometric filters can function as near spectroscopic quality redshifts. During training, a non-linear mapping between features and redshift is learned by the algorithm. A validation set is also required, and like the training set, truth information required for the validation set. The validation set is used to examine the effectiveness of the trained models prediction compared to the ground truth. This is useful for fine tuning 'hyper parameters' in the model to obtain an optimally successful model. Finally, there is the test data—the data whose redshifts are to be predicted with the trained machine learning model given its photometry and photometric colors. As was the case for template based codes, there are possible scenarios where blending may interfere with accurately predicting photometric redshifts. For example, if a training set does not contain carefully chosen representative data of recognized blends with secure redshifts, the machine learning algorithm will not be trained to predict redshifts for recognized blends accurately. Additionally, for unrecognized blends, the 'true redshift' is not well defined, consequentially prohibiting training a model to predict redshifts for unrecognized blends.

In this chapter, we study the impact that galaxy blending has on the statistical performance of photometric redshifts. Understanding deviations induced by blending requires knowledge of the 'ground truth'. We therefore require a truth catalog. We will introduce the N-body simulation which is used to create a truth catalog for this study. Additionally, we select two different photometric redshift codes to estimate p(z)s for galaxies: a template based code and a machine learning based code. Both approaches have pros and cons, and comparing and contrasting the results of both codes potentially provide insight on how to mitigate systematics related to blending. Following this, we will discuss the method in which we identify unrecognized blends in our catalogs. Finally, we will lay out the statistical metrics used to evaluate the performance of photometric redshift estimation codes on unblended galaxies and blended galaxies.

3.2 Data

Quantifying the impact blending has on any science case with real observational data is not possible, since with out knowing the ground truth, one cannot identify and disentangle unrecognized blends. To accommodate the need to know the ground truth, collaborations often turn to complex cosmological simulations that mimic intended surveys (DeRose et al., 2019). For our needs, we turn to Cosmological Data Challenge 2 (CosmoDC2) truth catalog (LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021).

CosmoDC2 is a large synthetic galaxy catalog designed to emulate LSST data quality. This allows the scientific community to prepare for LSST first light and beyond by exploring survey specific systematics and preparing analysis methods and pipelines ahead of time. The catalog is complete down to an r band magnitude of 28, occupies 440 deg^2 of sky area, with sources up to redshift $z = 3$. The catalogs provide a plethora of data quantities for each galaxy in the N-body simulation, including broadband filter magnitudes in $ugriz$ and true redshift among many others.

Additionally, high fidelity images of galaxies in the catalogs are generated, using these galaxy catalogs as input in a process called *ImSim*. The image simulations are highly realistic, mimicking several survey properties. Among them are survey observation strategy, physics of the atmosphere, telescope and detector, transient objects like variable Active Galactic Nuclei (AGN), galactic dust extinction, and stellar proper motions. The *ImSim* process used in CosmoDC2 simulates 5 years of LSST imaging.

We do not use the entire truth catalog or entire *ImSim* data set for our analysis. Instead, we restrict our attention to a subset of the full CosmoDC2 data catalog, covering 17 contiguous Nside=32 Hierarchical Equal Area isoLatitude Pixelation (HEALPix) [1] pixels of the truth catalog for a total area of $\approx$ 58 sq. deg.

The Vera Rubin Observatory Science Pipelines are used to reduce the *ImSim* data and process it end-to-end, producing a 5 year depth coadd catalog. This includes production of detection catalogs on coadded frames, deblending of recognized blends, and data products that include magnitudes,

[1] https://healpix.sourceforge.io/

positions, galaxy shapes for weak lensing, and processing flags, among many other quantities.

Ultimately we are left with two catalogs: a truth catalog and a coadd catalog. Although the Vera Rubin Science Pipelines attempt to deblend detected objects, the insidious case of unrecognized blends willremain in the coadd data. Having access to the ground truth, however, empowers us to empirically identify the occurrence of unrecognized blends. This will be fully discussed in 3.4.

As mentioned above, photo-z codes require training and validation data. Note the training and validation data are distinct from the ≈ 58 sq. degree data mentioned above that is being reserved for analysis. For our training and validation data, we first randomly select 5 well separated tracts—rectangular regions on the sky that are approximately 1.68 degrees on a side—from the coadd catalog to draw our samples from. The training sample is made up of approximately 100 thousand galaxies, while the validation sample is made up of approximately 200 thousand galaxies. At this point, we have *observed* properties of the galaxies for our validation and training sets, since they are selected from the coadd catalog. However, to validate the results of our training process, we require *truth* information—namely the true redshifts. To obtain the true redshifts, we perform a positional match with the coadd catalog and the truth catalog. Truth information can then readily be accessed. Dereddening of the training and validation galaxies is carried out using the Schlegel, Finkbinder and Davis maps (Schlegel et al., 1998). The training sample is used to train the prior for our template based code as well as our machine learning algorithm, while the validation sample is reserved to perform testing.

3.3 Photometric Redshifts Algorithms

3.3.1 Bayesian Photometric Redshift

Bayesian Photometric Redshift (BPZ) is a template based approach to photometric redshifts. BPZ evaluates the likelihood of a galaxy's observed colors given a set of SED templates at a given redshift. Using a Bayesian prior, $p(z, T|m_0)$ over apparent magnitude m_0 and template T, the postirior $p(z|C, m_0)$ is calculated by marginalizing over the different templates. The BPZ prior is constructed by multiplying two components together, $p(T|m_0)p(z|T, m_0)$ where the first component is a proportion that is a function of apparent magnitude, and the second is a prior over the expected redshift range as a function of apparent magnitude and template (Benítez, 2000).

To construct the templates, we first begin with the 190 template SEDs from Brown et al. (2014) in addition to several BC03 model SEDs. The models are expanded by applying two dust models that are extracted from two archetype galaxies which are drawn from cosmoDC2's color region, giving a grid of several reddenings. Subsequently, we run a Principal Component Analysis (PCA) on the template set to obtain 10 basis templates by keeping the first 10 eigenvectors. Following this, we use a Gaussian process to fit for the best-fit eigenvalue coefficients for a training set of cosmoDC2 galaxies using rest frame $ugrizy$ filters. The Gaussian sampling process step did not sample fluxes bluer than the u band. As a fix, we extend the SEDs by using simple tophats to constrain the bluer part of the SED. We then run K-means on the set of coefficients to get a spanning set of 229 templates (Kalmbach & Connolly, 2017). If a source was undetected in a photometric band, the non-detection was replaced with the 1σ detection threshold based on the signal to noise-based error forecast for 5 year LSST data (Ivezić et al., 2019). In principal, the error could also be estimated from the data directly. We choose to use the noise-based error forecast, however, because we had access to existing pipelines which employ the error forecast that has been used in previous photo-z analysis (Schmidt et al., 2020).

3.3.2 Trees for photo-z

Trees for Photo-z (TPZ) is a machine learning approach to estimating photo-zs. TPZ creates a random forest of prediction trees to estimate photo-z PDFs (Breiman et al, 1984). The training data is split iteratively at nodes into one of two branch paths, until terminating at a leaf. Splitting at the nodes is done such that the information gain among a bootstrapped sub-sample of the training data is maximized (Carrasco Kind & Brunner, 2013).

As was the case for BPZ above, non-detections are replaced with the 1σ detection threshold for the appropriate filter (Ivezić et al., 2019). We grow 125 trees to a minimum leaf size of 5. The features for the training, validation, and test set include $ugrizy$ magnitudes, their associated magnitude errors, and $u-g, g-r, r-i, i-z, z-y$ colors. Our redshift space is made of 200 bins, with redshift $0.001 > z > 3.0$. TPZ uses a smoothing scale parameter to smooth the individual photo-z PDFs when constructing them. To find an optimal smoothing scale, we perform a search through a variety of smoothing scales, producing photo-z PDFs for our validation set in each case. The optimal scale is selected based on the predicted PDFs which produce a Probability Integral

Transform (PIT) histogram closest to the uniform case, as measured by the 2 sample Kolmogorov-Smirnov (KS) test. As we will show in 3.5.1, the PIT distribution will be uniform in the case that the photo-z PDFs are robust estimators of the underlying n(z). In our case, the optimal smoothing scale was 1.75. Note, that according to the TPZ documentation, the width of the Gaussian filter used the smooth the individual PDFs is the smoothing scale multiplied by the resolution. In our case this leads to a Gaussian filter with width equal to ≈ 0.028.

3.4 Friends of Friends Matching

To identify blends, we utilize Friends of Friends (FoF) matching to identify 'groups'-collections of galaxies whose members can be connected along links no more than 1 arcsecond in length. We define an n-m FoF group as having n objects from the truth catalog and m galaxies from the coadd catalog being linked together. For example, a 1-1 group has one observed object linked to one truth object-the two being separated by no more than 1 arcsecond-which we may regard as a 'clean match', where we have successfully identified the truth objects counterpart in the simulated images. A 2-1 match is a group which contains two truth objects which can be linked with one object from the observed coadd catalog object. This is a potential case of blending, where the second truth galaxy was not identified because it formed an unrecognized blend with its neighbor in the truth catalog. These two cases will be the focus of our attention. The 1-1 case will form a baseline case for which we will measure statistical properties to quantify photo-z performance. We will then compare these to the metrics evaluated on 2-1 cases-instances of blends. Additionally, we show that the 2-1 case is by far the most common case of blending. This can be seen in Figure 3.4.1, where we display a 2 dimensional histogram detailing the number of various kinds of FoF groups. The bottom row displays instances of non-detections, while the leftmost column shows instances of false detections. Cases of n-m groups where n is greater than m represent unrecognized blends, e.g. 3-1 corresponds to a case where three galaxies have blended together into one detected galaxy. There are approximately 564,000 occurrences of the 2-1 group, the most occurrences of any blending group and thus the most common kind of blend. The next highest occurrence is for 3-1 groups, of which there are approximately 162,000.

Returning our attention to 2-1 groups, the distribution of magnitude separations between 2-1

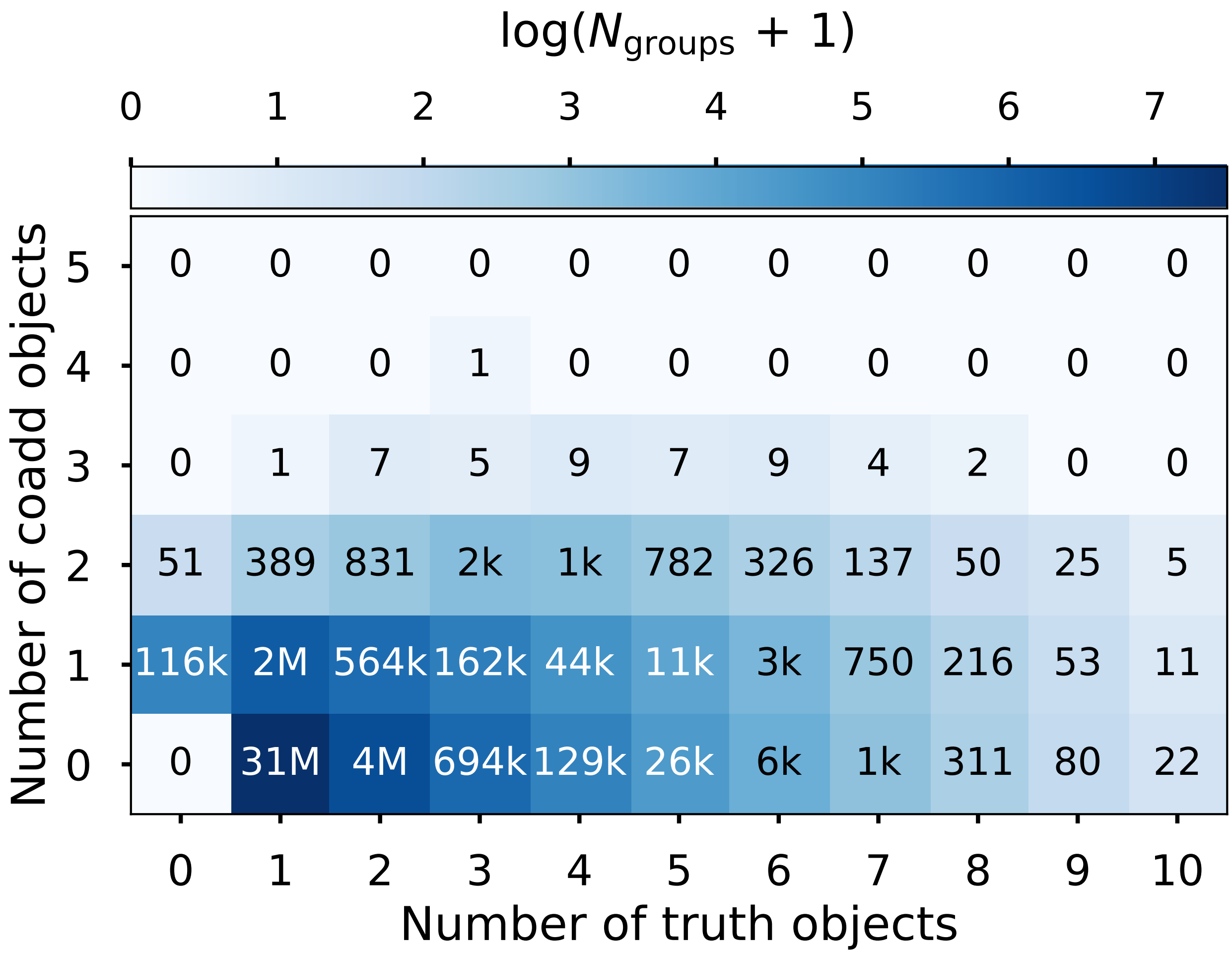

Figure 3.4.1: A two dimensional histogram showing the number of specific Friends of Friends groups. For example, when number of truth objects and number of coadd objects both equal one, we form the 1-1 group, made of clean matches, and have approximately 2 million such 1-1 groups. Also of interest is when the number of truth objects equals 2 and the number of coadd objects equals one, forming a 2-1 group. We interpret this as an unrecognized blend, and have approximately half a million such 2-1 groups. For context, the most common cases are 1-0 and 2-0 cases. These correspond to non-detections from the faint $\approx$ 28th magnitude galaxies which fall below the detection threshold in isolation in the case of the 1-0 group, and even when blended together in the 2-0 group.

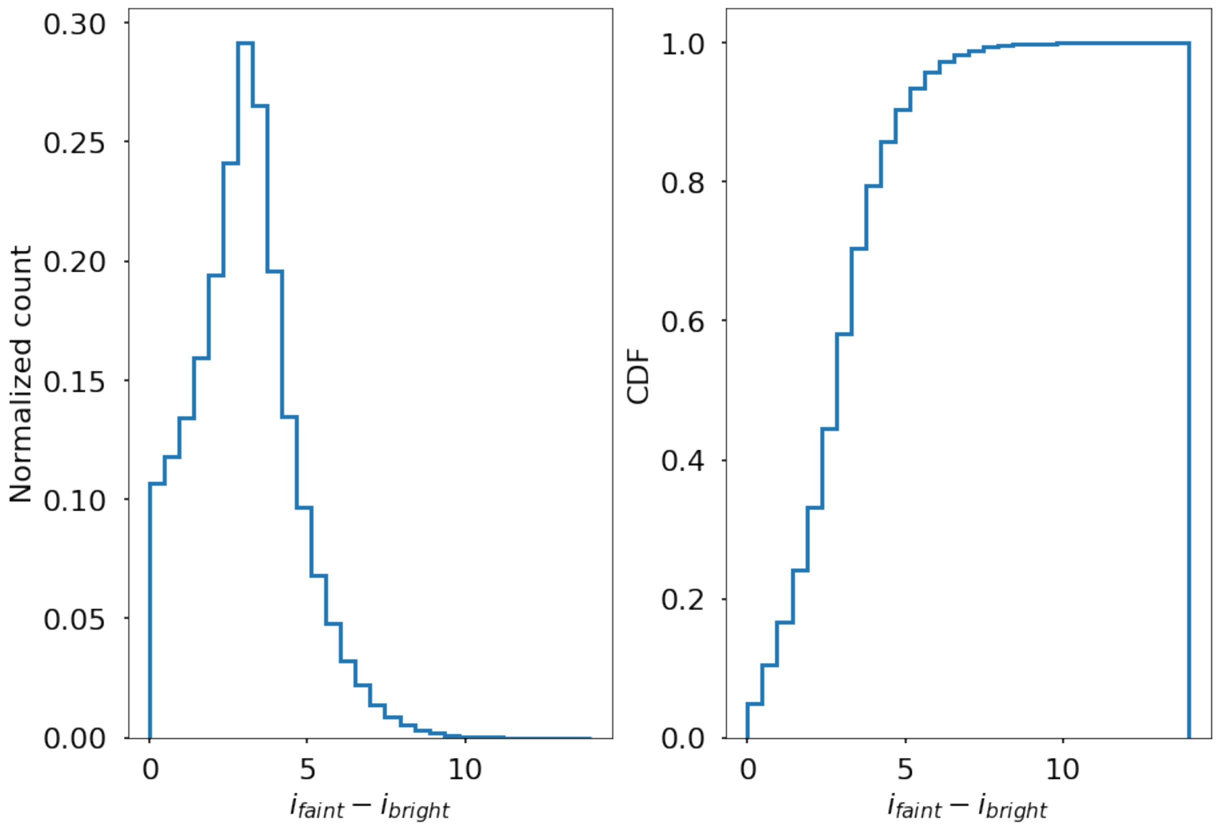

Figure 3.4.2: *left*: A histogram showing the distribution of the difference in magnitudes between members of the 2-1 group. The distribution peaks at a separation of approximately 3 magnitudes. *right*: the cumulative distribution function (CDF) of the histogram presented on the left. Only $\approx$ 10% of the 2-1 groups are separated by 1 magnitude or less. Approximately 50% of the 2-1 groups are separated by 3 magnitudes or less, with the remaining galaxies being separated by more than 3 magnitudes.

members is shown in the left panel of Figure 3.4.2. The distribution peaks at 3 magnitudes in the i band, and has an extended tail to the right going up to 14 magnitudes. The Cumulative Distribution Function (CDF) of the distribution is shown in the right panel. Approximately 50% of galaxies are separated by more than 3 magnitudes. An investigation shows that the change in color induced by blending together galaxies which are separated by three or more magnitudes is comparable to the photometric error associated with the galaxies.

3.5 Metrics Used To Evaluate Performance

3.5.1 Probability Integral Transform

The Probability Integral Transform (PIT) for a single galaxy is its Cumulative Distribution Function (CDF) evaluated at its true redshift. One can calculate the PIT for an ensemble of galaxies to assess the robustness of the ensembles photo-zs. If the photo-zs are true robust estimators of the true underlying redshift distribution, one expects the distribution of PIT values to form a uniform distribution, $U(0,1)$ (Bordoloi, Lilly, & Amara, 2010; Polsterer, D'Isanto, & Gieseke, 2016; Tanaka et al., 2018). Excursions from uniformity grant insight into shortcomings of the p(z)s. For example, build up at the edges of the distribution can be due to overly narrow PDFs, or excess catastrophic outliers. PDFs that are overly broad will create a PIT distribution that 'piles up' in the middle of the distribution, at the cost of a dearth of PIT values at the edges near 0 and 1.

3.5.2 Kolmogorov Smirnov Test for PITs and PDFs

The 2 sample Kolmogorov-Smirnov (KS) test compares the CDFs of two distributions to determine if they are drawn from the same distribution. The KS statistic is the maximum difference between the CDFs of the empirical functions.

$$KS = \max\big(|\text{CDF}(\hat{f}, z) - \text{CDF}(\widetilde{f}, z)|\big)$$

We apply the KS test in two scenarios in our work. First, we use the KS test to compare the stacked p(z)s for the ensemble sample to the true underlying n(z), to probe how accurately the photo-zs represent the redshift distribution. Second, the KS test is applied to the ensemble PIT histogram and the uniform distribution to quantify how robust the photo-zs are.

3.5.3 Point Statistics: Scatter, bias, and outlier rate

Calculating the point statistics of interest first requires we calculate $e_z = (z_{point} - z_{spec})/(1 + z_{spec})$ as an intermediate quantity. To calculate the scatter in a manner that is robust to outliers, we rely on the Interquartile Range (IQR), the difference between the 75th and 25th percentiles of the distribution of e_z. We then use the definition $\sigma_{IQR} = \text{IQR}/1.349$ to ensure that the area within

σ_{IQR} is the same as that within one standard deviation from a standard Normal distribution. We also calculate the bias, b_z, using the median of e_z in lieu of the mean to protect against outliers. Finally, we calculate the catastrophic outlier rate as the fraction of e_z values greater than 0.15.

3.5.4 Kullback-Leibler Divergence

The Kullback-Leibler Divergence (KL Divergence) is a measure of the relative differences between a probability distribution, $P(x)$, and a reference probability distribution $Q(x)$. Formally, it is calculated as

$$KLD = \sum_i P(x_i) \log\left(\frac{P(x_i)}{Q(x_i)}\right)$$

As an example, if $P(x)$ and $Q(x)$ are identical for all x, the KL divergence would be zero. In our study, we will compare the PIT histograms to the idealized case of a uniform distribution. That is to say, we will let $P(x) = \text{PIT}$ and $Q(x) = U(0,1)$-the uniform distribution between zero and one. Calculating the KL Divergence between these two distributions will give us another quantitative way to describe the departures of the PIT histograms from the idealized case of the uniform distribution.

3.6 Results

To gain some intuition for the impact of blending on photometric redshifts, we begin by examining point estimates. In Figure 3.6.1 we present four two dimensional histograms, each showing the BPZ mode versus true redshift. The bottom left panel displays results for our 1-1 FoF group-the 'clean' matches between the truth catalog and observed coadd catalog. Above this plot in the top right corner we show results for the 2-1 bright FoF galaxies-where blending has occurred and we find only one galaxy in our observed coadd catalog in spite of the fact that the truth catalog contains two galaxies at that position. Here we select the truth information of the brighter of the two blend galaxies to plot the true redshift. At first glance, there is not a tremendous difference between the 2-1 bright and 1-1 plots. This seemingly suggests that BPZ is primarily picking up on the photometry of the brighter blend galaxy when producing point estimates. This is supported by the third panel in the Figure: in the top right panel we plot the 2-1 faint galaxies-galaxies belonging to a 2-1 blend which are the fainter of the two blend galaxies. Two features stand out. There is small a population of galaxies along the $z_{mode} = z_{true}$ line, and otherwise the plot is essentially filled up with scattered points which have very little correlation between the BPZ mode and true redshift.

Taken together with the other plots, this suggests for most blends, since the brighter galaxy's photometry dominates over the fainter galaxy's photometry, the photometric redshift algorithm is primarily influenced by the brighter galaxy. Further investigation into the truth catalog revealed that the small population of 2-1 faint galaxies that seem to have accurate redshifts in fact occupy the same dark matter halos as their blend companions, making their redshifts essentially identical to that of their blend companions. Because these faint galaxies share the same true redshift as their brighter counterparts, and because the point estimate is primarily determined by the brighter galaxy, these faint galaxies *appear* to have accurate redshift point estimates. Put another way, the feature of galaxies along the $z_{mode} = z_{true}$ line in the 2-1 faint panel is due entirely to physical clustering of galaxies, not due to BPZ assigning accurate redshifts based on the fainter galaxy's flux.

Finally, in the bottom right panel of Figure 3.6.1 we select blends where we have enforced the magnitude separation between the blend galaxies to be 1 or fewer, and choose the truth information of the brighter galaxy to plot the true redshift. The scatter in this plot is larger than the 2-1 bright and 1-1 case. This illustrates the point that photo-z accuracy will be a function of magnitude separation of the two blends.

In Figure 3.6.2, we repeat the procedure outlined above plotting photo-z mode against true redshift, only now we use TPZ results instead of BPZ results. The same themes emerge in this Figure as the previous one. Comparing the plots here to Figure 3.6.1, we see fewer catastrophic outliers, and an over all tighter relationship between the mode and true redshift. It is worth noting that these results may be optimistic. While our training sample is under the recommended $\approx$ 300,000 training galaxies that will be targeted for Stage IV experiments (Newman et al., 2015), we have provided TPZ, a machine learning algorithm, with perfectly representative training data. This is likely not to be the case in real data, where the selection functions of spectroscopic surveys and imaging surveys differ considerably. On the other hand, our training set used in our study is smaller than what is recommended for Stage IV experiments like LSST.

We now turn our attention away from point estimates, and towards the full $p(z)$s. It is typical to stack the individual photo-z PDFs to make an ensemble PDF that is an estimate for the entire true

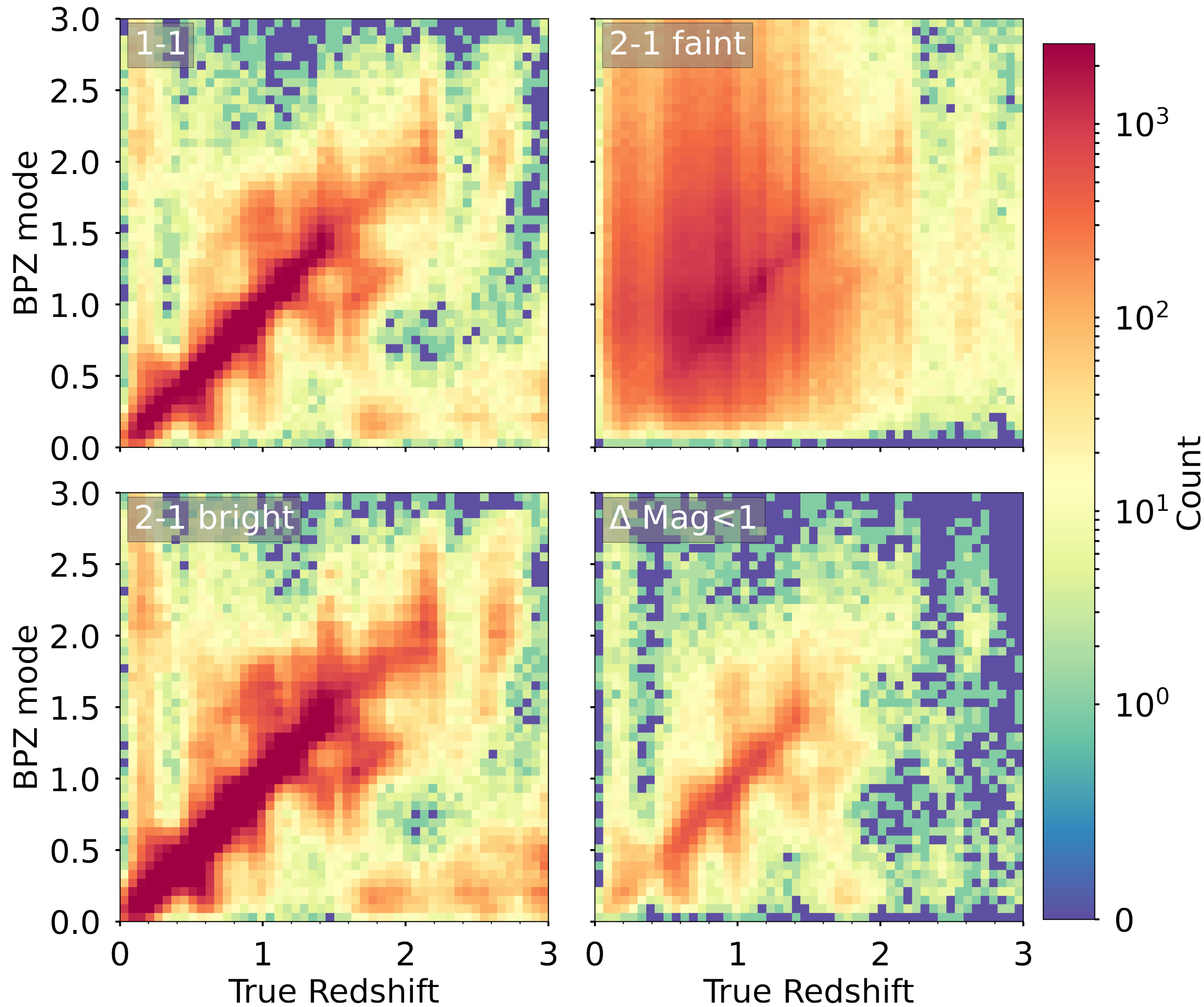

Figure 3.6.1: We show two dimensional histograms for true redshift versus the BPZ mode point estimate for four cases. Starting from the bottom left, going clockwise we show the results for the 1-1 case, the 2-1 case where we use the redshift of the brighter of the blended galaxies as the truth, the 2-1 case where we use the redshift of the fainter of the blended galaxies, and finally the 2-1 case where blended galaxies are separated by 1 magnitude or less, using the truth information from the brighter galaxy. We can see the 1-1 and 2-1 brighter panels look very similar, suggesting that blending is a minor effect when considering all blends. This is because $\approx 90\%$ of 2-1 blends have a magnitude separation 1 magnitude or less, suggesting the more common case of a wider magnitude separation leads to a case where the flux of the brighter member dominates over that of its fainter companion. When we enforce a magnitude separation of 1 or fewer between blends, as shown in the bottom right, we can see higher scatter, suggesting that when galaxies are of comparable magnitude, the color changes induced by blending have a non-negligible effect on point estimates.

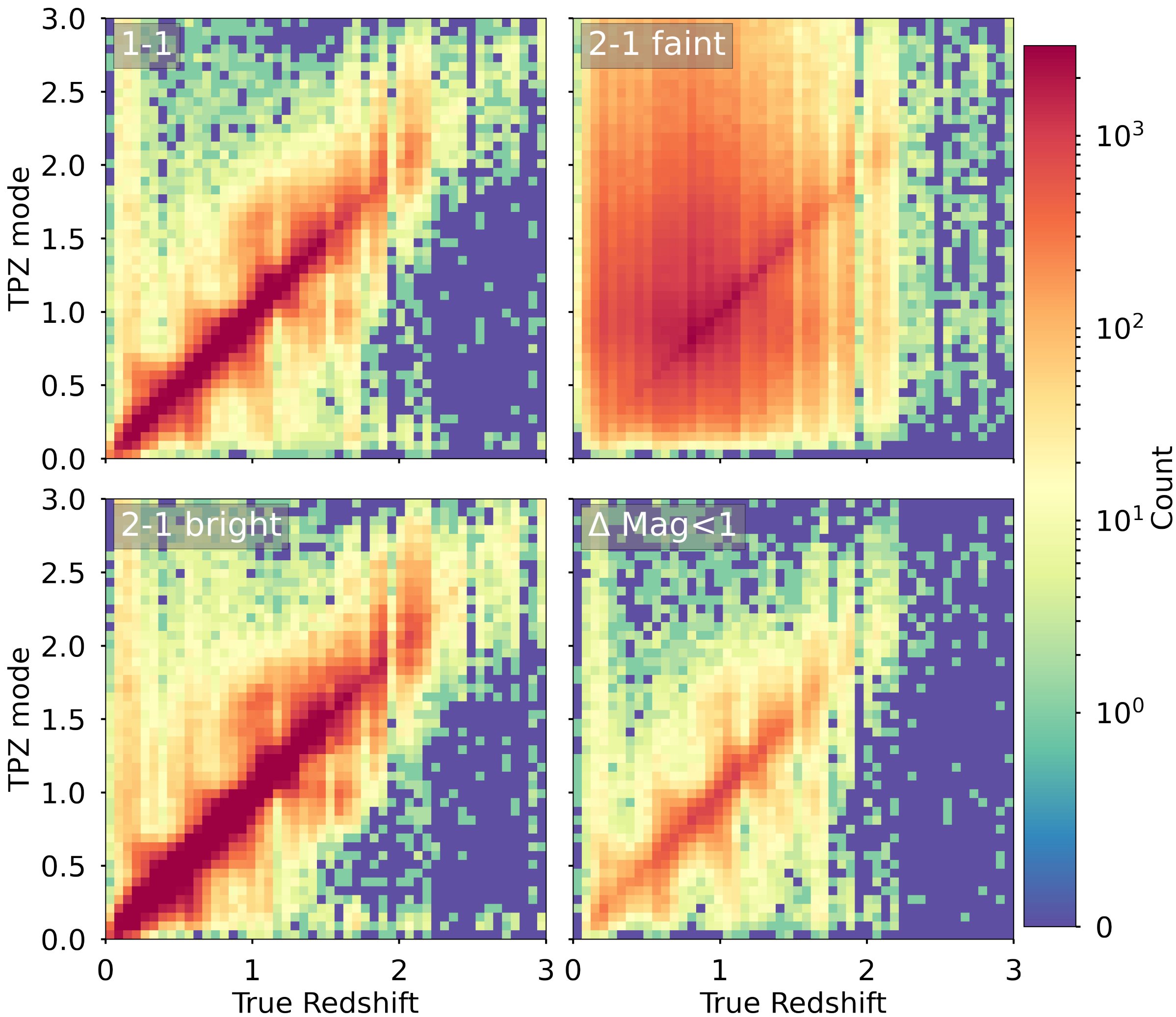

Figure 3.6.2: Same as Figure 3.6.1 but now with TPZ results. While the panels in this plot resemble those in the complimentary BPZ figure, we see fewer catastrophic outliers, and an over all tighter relationship between the mode and true redshift. In particular, the left panels in Figure 3.6.1 contain a locus of catastrophic outliers at a true redshift of $z \approx 0.02$. The left panels here also contain a locus of catastrophic outliers at this location, though fewer. Additionally, the BPZ figure has a collection of catastrophic outliers where the BPZ mean is ≈ 0.03 and true redshift $1.5 \geq z \geq 3$. These are not present in the left panels in this Figure.

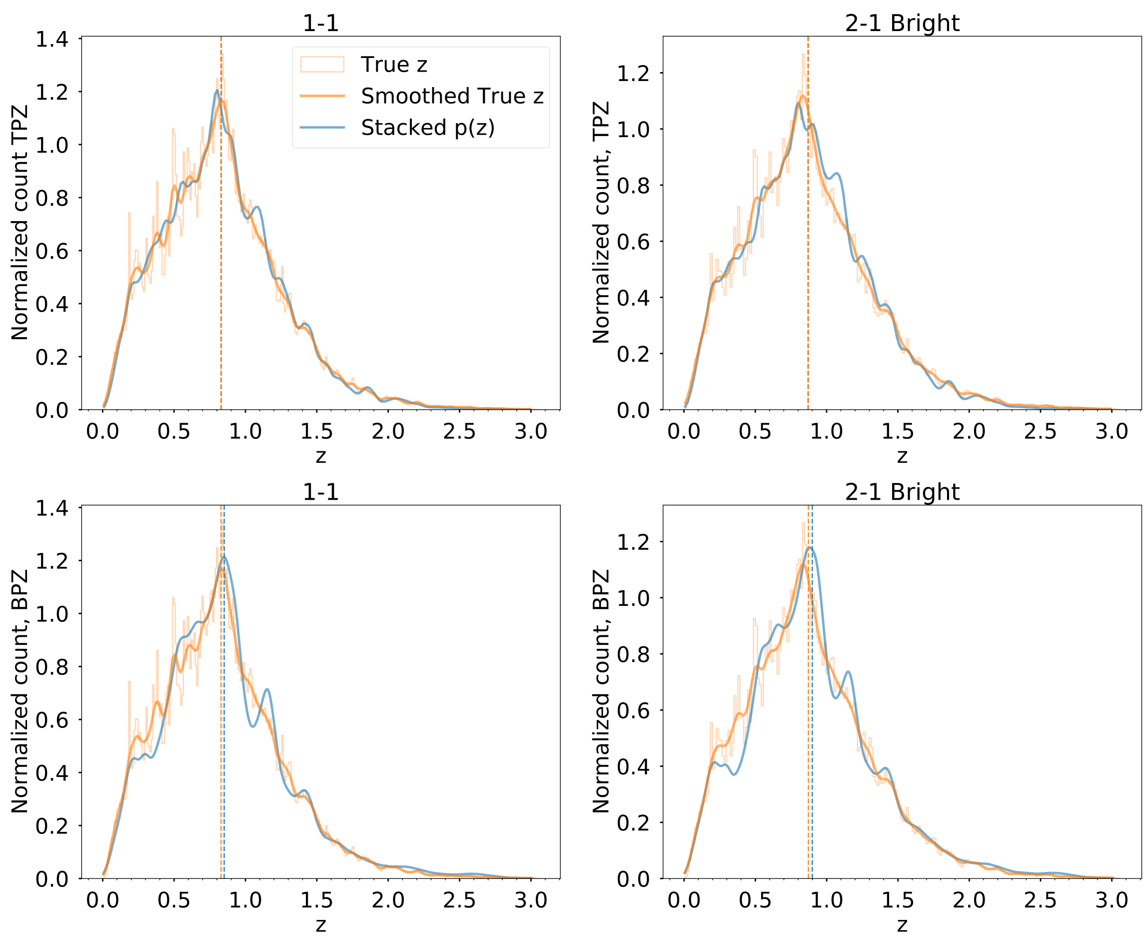

Figure 3.6.3: Each panel displays stacked $p(z)$ in blue, the true underlying $n(z)$ in orange histograms, and the smoothed true $n(z)$ in orange, obtained by convolving the orange histogram with a Gaussian filter. The top row displays results for our TPZ results, while the bottom row displays BPZ results. Also shown are vertical dashed lines which indicate the mean of the stacked $p(z)$ in blue, and the mean of the true $n(z)$ in orange. The panels show the 1-1 case (left column), and the 2-1 case with the brighter of the two blend galaxies selected to be the truth (right column). In the top row for TPZ the means of the stacked $p(z)$ and true $n(z)$ are within $\approx .02$ of each other for both cases. The mean of the stacked $p(z)$ moves from 0.8302 to 0.8703 going from the 1-1 case to the 2-1 case. For BPZ, the mean moves from 0.830 from the 1-1 case to 0.870 for the 2-1 case.

$n(z)$ distribution (Polsterer, D'Isanto, & Gieseke, 2016). One can then use statistical tests, such as those outlined in Section 3.5, to compare the true $n(z)$ to the stacked ensemble PDF to evaluate photo-z performance (Schmidt et al., 2020). In Figure 3.6.3 we present the results of photo-z PDF stacking for TPZ (top row) and BPZ (bottom row). The left column includes results for the 1-1 case for both codes, while the right column shows results for the 2-1 bright case for both codes. In each panel we present an orange histogram representing the true redshift distribution, a solid orange curve indicating the result of smoothing the histogram with a Gaussian filter, and a solid blue curve representing the stacked photo-z PDFs. Also shown are orange vertical dashed lines which indicate the true $n(z)$ mean and blue vertical dashed lines indicating the mean of the stacked photo-zs. Comparing the left column to the right column, both codes perform approximately as well on 2-1 blends as on single objects. While not catastrophic, there is *an* effect nonetheless. In Table 3.1 we show the KS test between the stacked $p(z)$s and the true $n(z)$. For BPZ, the KS statistics for the 1-1 case and 2-1 case are 0.025 and 0.033 respectively. For TPZ, the statistics are 0.007 and 0.013 respectively. For both BPZ and TPZ, we see a greater difference between the PDFs and $n(z)$ when we have blending.

To further quantify the difference in performance, we can measure the first three moments of the true redshift distribution and stacked PDFs, and compare the two. The means, standard deviations, and skew of the distributions are reported in Table 3.2. This Table allows us to examine the accuracy of the ensemble stacked $p(z)$ relative to the true $n(z)$ quantitatively. We may use the 1-1 FoF group as a baseline from which to relatively compare against the 2-1 bright case. The Δz between the true and estimated mean for the BPZ 1-1 case is 0.022, and 0.027 for the 2-1 bright case.

To provide some context for these numbers, we turn to the LSST Dark Energy (DESC) Science Requirement Document (SRD) (The LSST Dark Energy Science Collaboration et al., 2018). The SRD quantifies data requirements for cosmological probes if they are to enable an analysis of dark energy consistent with the Dark Energy Task Force definition of a Stage IV dark energy experiment. Photometric redshifts are a data product required for cosmological probes, and their requirements are outlined in the SRD. As an example, for the cosmic shear Year 1 analysis, the error on the mean photometric redshifts of each tomographic bin can be no more than $0.002(1 + z)$. This threshold evaluates to 0.0017 for our sample. Similarly to the redshift mean, the SRD places constraints

Blend Type	KS-PDF	KS-PIT	σ_{IQR}	Bias	Outlier Rate	KL Divergence
			BPZ			
1-1	0.025	0.143	0.045	0.006	0.086	0.258
2-1 Bright	0.033	0.175	0.052	0.008	0.120	0.324
			TPZ			
1-1	0.007	0.032	0.031	0.004	0.043	0.037
2-1 Bright	0.013	0.039	0.037	0.003	0.069	0.055

Table 3.1: We show the resulting values of our metrics, as measured on BPZ data (top two rows) and TPZ data (bottom two rows) for isolated galaxies and blended galaxies. Blending consistently worsens the performance of both codes across all metrics.

on the σ_z: for Year 1 cosmic shear analysis the error can be no greater than $0.006(1+z)$, which evaluates to 0.0051 for our sample.

Our above calculations show the error on the mean redshift, 0.022, is well beyond this threshold for using BPZ alone for the 1-1 case. For BPZ the difference between the true $n(z)$ σ_z and the $p(z)$ σ_z for the 1-1 case is 0.038. Reducing the error on the mean and σ_z will require additional analysis steps in addition to using a BPZ algorithm alone, such as jointly using angular clustering information (Matthews & Newman, 2012).

Turning our attention now to the TPZ data, we find the error in the mean redshift for the 1-1 and 2-1 bright case is 0.001 and 0.002, respectively. TPZ alone for the 1-1 case is narrowly within the SRD requirements for the mean. For 2-1 cases where we have blending, however, the error on the mean is outside of compliance. For the TPZ results σ_z=0.017—the SRD metrics are not met. Blending deteriorates this metric: for the 2-1 case σ_z=0.032, an overall change of 0.015. If blending is an additive systematic, it then follows that it is a cause for concern for Stage IV Dark Energy experiments. That is to say, even if it were possible to reduce the error on σ_z to make it compliant with the SRD, the effect of blending is large enough to take it out of compliance.

Beyond noting that blending can take photo-z means and widths out of compliance for the DESC SRD, we can quantify the degradation in error on the dark energy equation of state parameter, w. In Huterer et al. (2006) the authors investigate the impact of systematic errors on planned weak lensing surveys and compute the requirements on their contributions so that they are not a dominant source of the cosmological parameter error budget. Among the sources of systematic error are additive and multiplicative errors in photometric redshifts. The authors present degradation in the cosmological parameter accuracies as a function of our prior knowledge of the shift in redshift mean, $\Delta z = z_{phot} - z_{spec}$. In our study, for BPZ, $\Delta z = 0.022$ for the 1-1 case and 0.027 for the 2-1 case. The additional $\Delta z = 0.005$ shift corresponds to a degradation in w of approximately 160% according to Huterer et al. (2006). For TPZ, we measure $\Delta z = 0.001$ for the 1-1 case and 0.002 for the 2-1 case. The shift in Δz by 0.001 corresponds of approximately 25% in w.

Additionally, as a toy example we consider the change in the mean angular diameter distance, D_A, which corresponds to the shift in redshift mean induced by blending. We then calculate the change in the dark energy equation of state, w, which mimics such a change. We begin by assuming a fiducial flat ΛCDM *Planck* cosmology. For BPZ, the $p(z)$ mean goes from 0.851 for the 1-1 case to 0.899 for the 2-1 case. For the fiducial cosmology chosen, this corresponds to a change in D_A of 26 Mpc. We then define a flat wCDM cosmology, choosing values for H_0 and Ω_m equal to those of the fiducial *Planck* cosmology, and iteratively pick values for w and calculate D_A for the 1-1 case until it is shifted by 26 Mpc. We find that w needs to be approximately -1.085 to move D_A of the unblended bin by 26 Mpc. Repeating the procedure for TPZ, we find a shift of 23 Mpc between the unblended and blended $p(z)$s, and a value of $w = -1.075$ is needed to mimic the shift in it's angular diameter distance. We note that these shifts in w are likely overestimates. For weak lensing the quantity of interest is the distance ratio between two bins, which may shift less.

Another metric we can use to evaluate how accurate of a representation the stacked PDFs are of the overall $n(z)$ is the PIT. To reiterate from above, the distribution of PIT values is expected to be a uniform distribution, $U(0, 1)$, and any deviations from uniformity can be interpreted as deviations of the stacked photo-z PDF shapes from their corresponding $n(z)$ shapes. In Figure 3.6.4 we present the PIT distributions of four different cases in four panels. The top row shows results for TPZ, while the bottom row shows results for BPZ. The left column shows data for our 1-1 case, while the right panels show the 2-1 bright case. PIT distributions are shown in blue histograms. The uniform distribution is also plotted as a horizontal dashed black line. Across all cases, we see 'pile up' at the edges of the histograms, due to catastrophic outliers. In the BPZ data, the 'pile up' is greater on the left hand side of the distribution, in the bin which includes $z = 0$, than the right hand side of the distribution, in the bin which includes $z = 1$. This is due to a bias: the ensemble of BPZ PDFs tend to sit to the right—at higher redshift—of the true redshifts. This feature appears both in the 1-1 case and the 2-1 bright case, indicating it is a feature inherent to BPZ PDFs, and not attributed to blending.

Blend Type	$p(z)$ Mean	$n(z)$ Mean	$p(z)$ Std.	$n(z)$ Std.	$p(z)$ Skew	$n(z)$ Skew
			BPZ			
1-1	0.851	0.829	0.912	0.874	1.171	1.084
2-1 Bright	0.899	0.872	1.009	0.969	1.340	1.261
			TPZ			
1-1	0.830	0.829	0.857	0.874	1.037	1.084
2-1 Bright	0.870	0.872	0.937	0.969	1.172	1.261

Table 3.2: We quantify the shapes of the $p(z)$ and $n(z)$ distributions for the 1-1 and 2-1 Bright case by using the first three moments of the distributions: the mean, standard deviation, and skew. The nth moment of a distribution $F(x)$is defined as $\int x^n F(x)\mathrm{d}x$. Shown are results for BPZ in the top two rows and TPZ in the bottom two rows. Note, the $p(z)$ statistics for the 2-1 Bright and 2-1 Faint cases are identical. One can compare the properties of the true underlying $n(z)$ to the properties of the $p(z)$.

Additionally, we can see the 2-1 bright PIT distribution is 'shallower' than the 1-1 case, with its distribution sitting slightly further below the uniform distribution. The effect of creating a 'shallower' PIT histogram from blending can be interpreted as having photo-z PDFs that are overly narrow, and thus overly confident. This 'shallower' quality is quantified in Table 3.1, where we show the KS test on the PIT histograms and calculate the KL divergence between the PIT histogram and uniform distribution. The 1-1 case for BPZ has a KS statistic of 0.143, while the 2-1 case has a higher statistic of 0.175, indicating a greater difference between the two distributions when blending occurs. This is supported by the KL divergence as well, where the 1-1 case has a KL divergence of 0.258 while the 2-1 case has a higher KL divergence 0f 0.324, indicating a greater degree of difference between the 2-1 PIT distribution and its corresponding uniform distribution.

For the TPZ data, we see a relatively minor bias to high values in the 1-1 case that is not in the 2-1 case. This is in contrast to the BPZ data, where there was a persistent bias in the blended and unblended cases. The 1-1 case for TPZ data has a 'hump' in the middle of its distribution, that gets flattened out in the 2-1 case. This indicates that the individual photo-z PDFs were overly broad in the 1-1 data, and slightly narrower by comparison in the 2-1 case. Like we saw with the BPZ data, blending here has the effect of suppressing the amplitude of the PIT histogram. Although the center of the PIT distribution is closer to the uniform case for the 2-1 case, the catastrophic outlier rate on the high and low end increases in the 2-1 case. Quantitatively, the KS statistics for the 1-1 and 2-1 case are .032 and .039. The KL divergence shows the same trend, as we see the statistic increases from 0.037 in the 1-1 case to 0.055 in the 2-1 case. These measurements show that overall blending deteriorates photo-z performance.

3.7 Conclusion

We utilize an extragalactic truth catalog to generate realistic image simulations of 5 year depth LSST like observations. Using the truth catalog, we identify unrecognized blends-cases of galaxy blending where the deblender failed to recognize distinct galaxies and did not attempt to deblend them. We identify a variety of different kinds of blends and their relative occurrences. We identify the case where two galaxies blend together to make one observed galaxy as the most common kind of blend, and refer to the system as a '2-1 blend'. Because they make up the majority of the blend cases, we choose to focus our attention for the remainder of the study on 2-1 blends, leaving studies of other cases to future work.

We run two photo-z codes on the observed data, both blended and non-blended galaxies, to understand the impact of galaxy blending on phtomoteric redshift estimation: BPZ, a template based approach, and TPZ, a machine learning approach. We define several statistical metrics to quantify the impact of blending. Some of our metrics are science-agnostic metrics—both established and emerging—to stress-test the ensemble properties of photo-z PDF catalogs. We observe that in every metric used in our study—while not catastrophic—blending nonetheless degrades photo-z performance. The degree to which these changes are tolerable will depend entirely on the science cases at hand.

The remaining metrics—the relative error in the first and second moment of the ensemble PDFs to their corresponding true redshift distribution—can be compared to science requirements for stage IV cosmology surveys. We find that the error due to blending is comparable to or larger than the error budget allotted to photo-z requirements for weak lensing probes.

To gain some intuition for the potential impact on measurement of cosmological parameters, we calculate the degradation in the equation of state parameters for dark energy, w, using previous work (Huterer et al., 2006). If used alone to calculate photometric redshifts, blending has the potential to degrade the accuracy of the dark energy equation of state parameter w by approximately 160% for BPZ. By comparison, for TPZ accuracy is degraded by approximately 25%.

There are some avenues that have displayed promise in mitigating the problems outlined above.

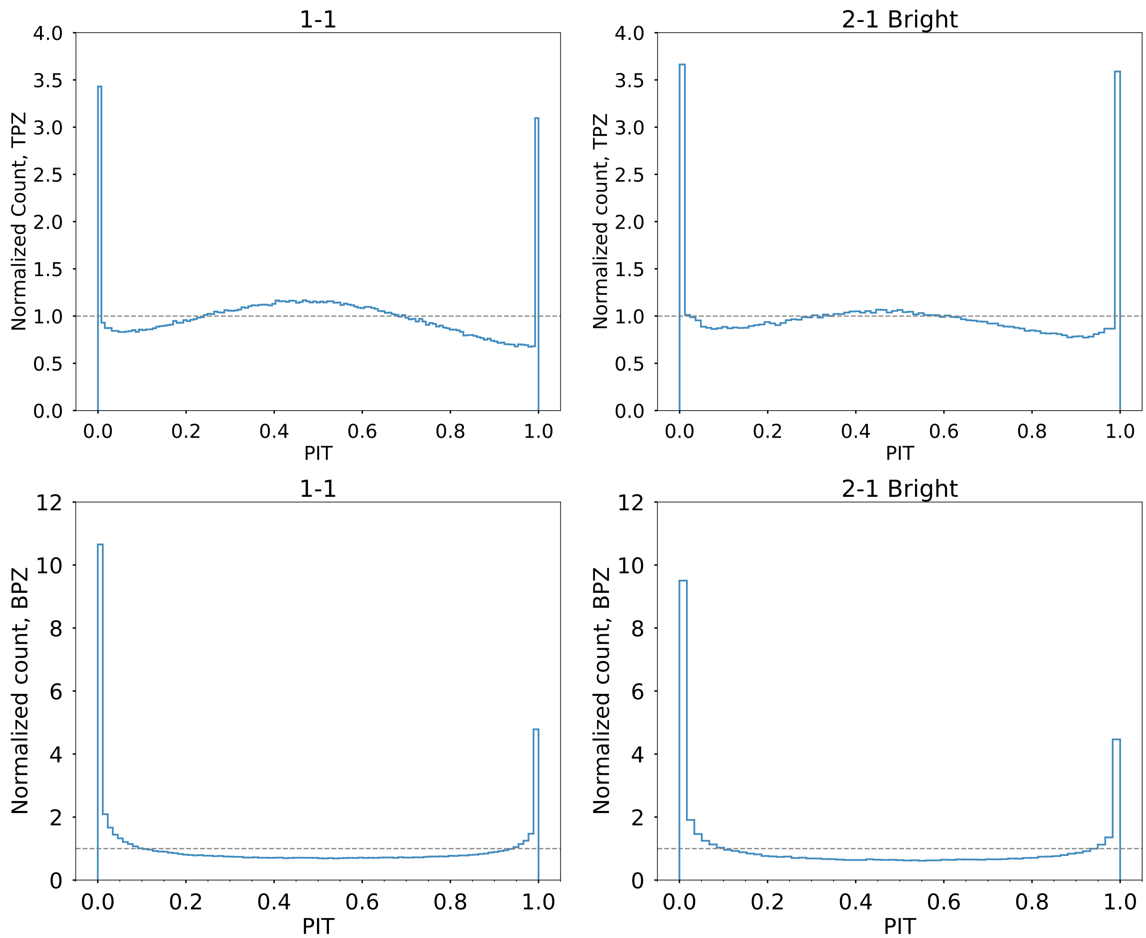

Figure 3.6.4: We show the Probability Integral Transform (PIT) histograms in blue in two rows of two columns. The top row displays results for our TPZ data, while the bottom row displays results for our BPZ data. Each panel also has a dashed black line, indicating the histogram of a uniform distribution. The panels show the 1-1 case (left column), the 2-1 case with the brighter of the two blend galaxies selected to be the truth (right-column). In an idealized case, the PIT distribution will match the uniform distribution if the photo-zs are robust estimators of the true underlying $n(z)$. In all panels, we see 'pile up' at the edges of the distributions due to outliers. For the TPZ data, the 1-1 case shows the $p(z)$ PDFs have a slight bias to high values due to the right most bin having slightly less height than the left most bin. The 2-1 bright case does not indicate any $p(z)$ bias. For the BPZ data we see an imbalance of pile up across all cases. The 1-1 and 2-1 bright case indicate the photo-z PDFs are biased high, while the 2-1 faint case is biased low. In the case of blending, our PIT histograms indicate that the photo-z algorithms studied here are primarily picking up on the brighter of the two galaxies, and performing similar to the 1-1 case.

They involve adding more complex approaches to photometric redshift estimation by jointly including ancillary information along with photometry. For example, Morrison et al. (2017) demonstrate that estimating the redshift distributions of photometric galaxies with unknown redshifts by spatially cross-correlating them against a reference sample with known redshifts can provide redshift powerful redshift constraints. Additionally, Myles et al. (2021) use a self organized map photo-z code—a machine learning method—jointly with clustering information, and shear ratios, which provide constraints on redshifts from the ratios of the galaxy-shear correlation functions at small scales, to provide tighter constraints on redshift distributions than using a photo-z code alone. Using the joint combination of methods, it may be possible to 'learn' and calibrate out the effect that blending has on photometric redshifts.

Chapter 4

Conclusion

We utilize a Rubin Observatory LSST precursor, the Deep Lens Survey, to investigate methods to mitigate errors on the $n(z)$ shape can impact cosmology constraints when using the combination of galaxy clustering, galaxy-galaxy lensing, and cosmic shear probes. To correct for the width of the $n(z)$, we use a validation set of data for which we have spectroscopic data, and empirically broaden individual photo-z $p(z)$ with a best-fit Gaussian filter. The width of the Gaussian filter is selected by minimizing the Kullback-Leibler divergence between the probability integral transform of the photo-z PDFs and the uniform distribution. We find that correcting the DLS $n(z)$ in this manner results in a shift in the constraint on the growth of structure parameter S_8 to lower values. The dependence of cosmology parameters on correct knowledge of the tomographic bin widths shown in this analysis confirms the stringent requirements on these parameters shown in forecasts for Stage IV surveys (Ma, Hu, & Huterer, 2006). While our result is not as constraining as other weak lensing measurements of S_8, we find agreement with them and *Planck* for our fiducial set of spatial scales for the joint combination of galaxy clustering and galaxy-galaxy lensing, and for cosmic shear, separately. In the joint combination of all three probes, we constrain S_8 to be $\approx 1\sigma$ lower than previous weak lensing studies. We also find the results are highly sensitive to the scale cuts used, for all probes. This confirms previous analysis of Deep Lens Survey data (Chang et al., 2019), and indicates there are lingering systematic effects that have not been adequately addressed.

We turn to realistic image simulations to understand the impact of galaxy blending on photometric redshifts. We employ the Hyper Suprime Cam science pipeline on the simulated data to perform detection, deblending, and photometry. We identify $n-m$ blends, where n is the number of truth objects and m is the number of detected galaxies. Enumerating the different kinds of blends in our analysis, we find that the 2-1 case dominates over all other cases of galaxy blending. We restrict our attention to 2-1 blends in the analysis. We run two different photo-z codes on our measured data: BPZ—a template based code—and TPZ—a machine learning based code. After outlining several science-agnostic metrics, we observe that blending degrades the performance of photo-z performance. The degree to which these degradations matter requires context, the different science cases come with different requirements. With an eye towards future weak lensing surveys, we examine the degradations in photo-z performance relative to the stringent requirements outlined in the LSST Science Requirement Document. We find that the error that blending induces on the photo-z mean and scatter exceeds the allowances when using BPZ alone, and that the errors for TPZ alone are on the order of the allowances. If blending is an additive systematic, it is a cause for concern for up coming weak lensing cosmology studies.

The coming generation of weak lensing cosmology surveys has the potential to characterize fundamental properties of dark energy, teasing out any possible time dependence in its equation of state parameter. Additionally, they can shed light on the S_8 tension: the $\approx 2 - 3\sigma$ difference between some weak lensing measurements of S_8 and the *Planck* constraints on S_8. This S_8 tension may be providing a hint of new physics beyond the standard ΛCDM cosmological model (Amendola et al., 2018). However, interpretation of these 2-3 σ tensions as a departure from the concordance model must be tempered by discretion. Weak lensing surveys suffer from a myriad of pernicious systematic effects, and the S_8 tension may be due entirely to these. High precision measurements demand robust mitigation of systematic errors as a necessary prerequisite if they are to address the S_8 tension, or test cosmological models more broadly. We have explored a mitigation strategy for correcting the overall shape of tomographic redshift distributions in real data. We leave a more detailed study using realistic simulations to future work, so we may examine the effect of shape of redshift distribution in isolation of other sources of error. We have also uncovered galaxy blending as a source of systematic error that threatens to take future weak lensing studies out of the required compliance. Redshift estimation may require more information in addition to photometry to provide the desired constraints, like clustering information and shear ratios. We leave the exploration of this topic to future work.

Bibliography

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, Phys.Rev. D,, 98, 043526

Albrecht A., et al., 2006, arXiv e-prints, astro–ph/0609591

Amendola L., Appleby S., Avgoustidis A., Bacon D., Baker T., Baldi M., Bartolo N., et al., 2018, LRR, 21, 2. doi:10.1007/s41114-017-0010-3

Benítez, N. 2000, ApJ, 536, 571

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Bordoloi R., Lilly S. J., Amara A., 2010, MNRAS, 406, 881. doi:10.1111/j.1365-2966.2010.16765.x

Bosch J., Armstrong R., Bickerton S., Furusawa H., Ikeda H., Koike M., Lupton R., et al., 2018, PASJ, 70, S5. doi:10.1093/pasj/psx080

Bradshaw A. K., Jee M. J., Tyson J. A., 2019, MNRAS, 482, 1190

Breiman L., Friedman J. H., Olshen R. A., Stone C. J., 1984, Classification and Regression Trees, Statis- tics/Probability Series. Wadsworth Publishing Company, Belmont, California, U.S.A.

Bridle S., King L., 2007, NJPh, 9, 444. doi:10.1088/1367-2630/9/12/444

Brown M. J. I., Moustakas J., Smith J.-D. T., da Cunha E., Jarrett T. H., Imanishi M., Armus L., et al., 2014, ApJS, 212, 18. doi:10.1088/0067-0049/212/2/18

Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483. doi:10.1093/mnras/stt574

Catelan P., Kamionkowski M., Blandford R. D., 2001, MNRAS, 320, L7. doi:10.1046/j.1365-8711.2001.04105.x

Chang C., Wang M., Dodelson S., Eifler T., Heymans C., Jarvis M., Jee M. J., et al., 2019, MNRAS, 482, 3696. doi:10.1093/mnras/sty2902

Choi, A., Tyson, J. A., Morrison, C. B., et al. 2012, ApJ, 759, 101

Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, AJ, 132, 926

Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8

Dawson W. A., Schneider M. D., Tyson J. A., Jee M. J., 2016, ApJ, 816, 11. doi:10.3847/0004-637X/816/1/11

DeRose J., Wechsler R. H., Becker M. R., Busha M. T., Rykoff E. S., MacCrann N., Erickson B., et al., 2019, arXiv, arXiv:1901.02401

Dekel, A., & Lahav, O. 1999, ApJ, 520, 24

Dodelson, S., 2003, *Modern Cosmology* Amsterdam (Netherlands): Academic Press.

Elvin-Poole, J., Crocce, M., Ross, A. J., et al. 2018, Phys.Rev. D,, 98, 042006

Geller, M. J., Dell'Antonio, I. P., Kurtz, M. J., et al. 2005, ApJ, 635, L125

Hamana T., Shirasaki M., Miyazaki S., Hikage C., Oguri M., More S., Armstrong R., et al., 2020, PASJ, 72, 16. doi:10.1093/pasj/psz138

Hartplap J. et al., 2007 A&A, 464, 399

Heymans C., Tröster T., Asgari M., Blake C., Hildebrandt H., Joachimi B., Kuijken K., et al., 2020, arXiv, arXiv:2007.15632

Hikage, C., Oguri, M., Hamana, T., et al. 2018, arXiv e-prints , arXiv:1809.09148.

Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2018, arXiv e-prints , arXiv:1812.06076.

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454.

Hildebrandt H., Arnouts S., Capak P., Moustakas L. A., Wolf C., Abdalla F. B., Assef R. J., et al., 2010, AA, 523, A31. doi:10.1051/0004-6361/201014885

Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, MNRAS, 478, 592

Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101. doi:10.1111/j.1365-2966.2005.09782.x

Huterer D., et al., 2015, APh, 63, 23

Ivezić Ž., Kahn S. M., Tyson J. A., Abel B., Acosta E., Allsman R., Alonso D., et al., 2019, ApJ, 873, 111. doi:10.3847/1538-4357/ab042c

Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338

Jarvis J. F., Tyson J. A., 1979, SPIE, 172, 422. doi:10.1117/12.957112

Jee, M. J., Tyson, J. A., Hilbert, S., et al. 2016, ApJ, 824, 77

Jee, M. J., Tyson, J. A., Schneider, M. D., et al. 2013, ApJ, 765, 74

Jee, M. J., & Tyson, J. A. 2011, PASP, 123, 596

Jee, M. J., Blakeslee, J. P., Sirianni, M., et al. 2007, PASP, 119, 1403

Kaiser N., 1984, ApJL, 284, L9. doi:10.1086/184341

Kalmbach J. B., Connolly A. J., 2017, AJ, 154, 277. doi:10.3847/1538-3881/aa9933

Kilbinger M., 2015, RPPh, 78, 086901

Kilbinger M., et al., 2013, MNRAS, 430, 2200

Kilbinger M., et al., 2009, A& A, 497, 677

Krause, E., Eifler, T. F., Zuntz, J., et al. 2017, arXiv:1706.09359

Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64

LSST Dark Energy Science Collaboration (LSST DESC), Abolfathi B., Alonso D., Armstrong R., Aubourg É., Awan H., Babuji Y. N., et al., 2021, ApJS, 253, 31. doi:10.3847/1538-4365/abd62c

LSST Dark Energy Science Collaboration, Abolfathi B., Alonso D., Armstrong R., Aubourg É., Awan H., Babuji Y. N., et al., 2020, arXiv, arXiv:2010.05926

The LSST Dark Energy Science Collaboration, Mandelbaum R., Eifler T., Hložek R., Collett T., Gawiser E., Scolnic D., et al., 2018, arXiv, arXiv:1809.01669

Ma Z., Hu W., Huterer D., 2006, ApJ, 636, 21. doi:10.1086/497068

MacCrann N., Becker M. R., McCullough J., Amon A., Gruen D., Jarvis M., Choi A., et al., 2020, arXiv, arXiv:2012.08567

Mandelbaum R., 2018, ARA&A, 56, 393

Matthews D. J., Newman J. A., 2012, ApJ, 745, 180

Melchior P., Moolekamp F., Jerdee M., Armstrong R., Sun A.-L., Bosch J., Lupton R., 2018, AC, 24, 129. doi:10.1016/j.ascom.2018.07.001

Molino A., Benítez N., Ascaso B., Coe D., Postman M., Jouvel S., Host O., et al., 2017, MNRAS, 470, 95. doi:10.1093/mnras/stx1243

Morrison, C. B., Scranton, R., Ménard, B., et al. 2012, MNRAS, 426, 2489

Morrison C. B., Hildebrandt H., Schmidt S. J., Baldry I. K., Bilicki M., Choi A., Erben T., et al., 2017, MNRAS, 467, 3576. doi:10.1093/mnras/stx342

Myles J., Alarcon A., Amon A., Sánchez C., Everett S., DeRose J., McCullough J., et al., 2021, MNRAS, 505, 4249. doi:10.1093/mnras/stab1515

Newman J. A., Abate A., Abdalla F. B., Allam S., Allen S. W., Ansari R., Bailey S., et al., 2015, APh, 63, 81. doi:10.1016/j.astropartphys.2014.06.007

Padmanabhan, N., Schlegel, D. J., Finkbeiner, D. P., et al. 2008, ApJ, 674, 1217

Pen, U.-L. 1998, ApJ, 504, 601

Planck Collaboration, Aghanim N., Akrami Y., Ashdown M., Aumont J., Baccigalupi C., Ballardini M., et al., 2020, A&A, 641, A6. doi:10.1051/0004-6361/201833910

Polsterer K. L., D'Isanto A., Gieseke F., 2016, arXiv, arXiv:1608.08016

Prat, J., Sánchez, C., Fang, Y., et al. 2018, Phys.Rev. D, 98, 042005

Rozo E., et al., 2016, MNRAS, 461, 1431

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schneider, P., 2006, *Extragalactic Astronomy and Cosmology: An Introduction* Berlin (Germany): Springer

Schneider, P., Kochaneck, C., & Wambsganss, J. 2006, *Gravitational Lensing: Strong, Weak and Micro* Berlin (Germany) : Springer

Schmidt S. J., Malz A. I., Soo J. Y. H., Almosallam I. A., Brescia M., Cavuoti S., Cohen-Tanugi J., et al., 2020, MNRAS, 499, 1587. doi:10.1093/mnras/staa2799

Schmidt, S. J., & Thorman, P. 2013, MNRAS, 431, 2766

Singh S., Mandelbaum R., Seljak U., Slosar A., Vazquez Gonzalez J., 2017, MNRAS, 471, 3827

Smith R. E., et al., 2003, MNRAS, 341, 1311

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152

Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, Publications of the Astronomical Society of Japan, 70, S9.

Troxel, M. A., MacCrann, N., Zuntz, J., et al. 2018, Phys.Rev. D, 98, 043528.

van Uitert, E., Joachimi, B., Joudaki, S., et al. 2018, MNRAS, 476, 4662.

Yoon, M., Jee, M. J., Tyson, J. A., et al. 2019, ApJ, 870, 111.

Wittman D., Bhaskar R., Tobin R., 2016, MNRAS, 457, 4005

Wittman, D., Ryan, R., & Thorman, P. 2012, MNRAS, 421, 2251

Wittman, D. M., Tyson, J. A., Dell'Antonio, I. P., et al. 2002, Proc. SPIE, 4836, 73

Xavier, H and Abdalla, F. B and Jochimi, B. 2016, MNRAS, 459, 3693

Zhan H., Tyson J. A., 2018, RPPh, 81, 66901

Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45