Towards High-redshift Cosmology with Lyman-break Galaxies Detected by LSST

John Franklin Crenshaw

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:
Andrew J. Connolly, Chair
Miguel F. Morales
Masha Baryakhtar

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

Towards High-redshift Cosmology
with Lyman-break Galaxies Detected by LSST

John Franklin Crenshaw

Chair of the Supervisory Committee:
Andrew J. Connolly
Department of Astronomy

The Vera C. Rubin Observatory is set to begin the Legacy Survey of Space and Time (LSST), a generation-defining astronomical survey that will image the entire southern sky in 6 photometric bands to unprecedented depth. LSST promises to discover hundreds-of-millions of high-redshift galaxies, opening a huge, previously unprobed volume of the universe to precision cosmology. These high-redshift constraints will provide new ways to test the standard cosmological model and have the potential to shed new light on the many tensions present in modern cosmology, including the evolution of dark energy, the sum of neutrino masses, and the mass density of the cosmos. Extracting information about the evolution of the high-redshift universe from LSST data will require careful modeling and new analysis tools to control systematic errors.

This dissertation develops new methods for estimating the distance to galaxies using photometric data (photometric redshifts, or photo-z's) and studying the systematic errors that plague them. Using machine learning tools, we show that galaxy spectral templates can be learned directly from broadband photometry, increasing the accuracy and precision of template-based photo-z estimation, which will figure prominently in the analysis of high-redshift galaxies. Using normalizing flows, we develop a statistical forward model of photometric galaxy catalogs, enabling new and more reliable studies of photo-z calibration,

including consistent evaluation of photo-z posterior distributions.

We then discuss optimizing the LSST survey strategy for the detection and photo-z estimation of high-redshift galaxies, before forecasting number densities by combining simulations of LSST with calibration data from precursor surveys. Using this model, we forecast the power of LSST for constraining the growth of large scale structure and the evolution of dark energy, finding that a joint analysis of high- and low-redshift data increases the constraining power of LSST by a factor of three compared to constraints from low-redshift data alone. We also study various sources of systematic error, quantifying their impact on cosmological constraints, and considering how data from LSST and CMB lensing surveys can be combined to reduce the impact of these errors.

A series of appendices present research on wave-front estimation for the Rubin Observatory's active optics system (AOS), which maintains the telescope's optical alignment and mirror figure to correct for optical aberrations and deliver high image quality across Rubin's wide field of view. First we describe the physical components of the AOS, before deriving an algorithm for wave-front estimation in Rubin's fast, wide-field optical system. We then introduce and validate a deep learning algorithm for wave-front estimation, showing it to be faster and more robust than traditional methods. We conclude by studying the information content carried by the shape and intensity of stars in the defocused images used for wave-front estimation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am very fortunate to say I have absolutely loved my time in graduate school, and this is in large part thanks to a huge number of collaborators, mentors, and friends who have shaped and enriched my graduate career.

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Andy Connolly, who has been an outstanding mentor — scientifically, professionally, and personally. His guidance has been instrumental in shaping the direction of my research, providing invaluable opportunities while also allowing me to follow my own interests. Beyond his mentorship, Andy has been incredibly fun to work with. Not many graduate students can genuinely say their weekly meeting with their advisor is the highlight of their week, but for the past six years that has consistently been true for me. Our meetings have always intermixed interesting science with plenty of laughs, and I always leave feeling more excited, focused, and empowered than when I arrived. I cannot thank Andy enough for his insight, encouragement, and unwavering good humor, and I look forward to a lifetime of continued collaboration and friendship.

I am also very grateful for the community provided by the UW's Dark Universe Science Center (DUSC). Professors Matt McQuinn, Miguel Morales, Marilena Loverde, and Masha Baryakhtar have each been extremely supportive and great mentors during my PhD. I would like to especially thank Masha for agreeing to join my reading committee only two weeks before my defense to fill a key vacancy. Fellow DUSC graduate students Murali Saravanan and Caio Nascimento have also been great companions throughout the latter half of my PhD. I have learned so much from both of them.

A huge fraction of my graduate research has taken place within the Dark Energy Science

Collaboration (DESC). DESC has provided a hugely supportive community that has dramatically broadened my opportunities and connections during graduate school. This community was especially important through the COVID-19 lock downs, during which, without DESC, the first few years of graduate school would have felt far more isolated and lonely. I would like to thank (in no particular order) Sam Schmidt, Alex Malz, Eric Charles, Eric Gawiser, Nora Shipp, Tanveer Karim, Claire-Alice Hébert, Alex Gagliano, Alex Broughton, Rebecca Chen, Theo Schutt, Biprateep Dey, Gautham Narayan, Boris Leistedt, Francois Lanusse, Renée Hložek, Tesla Jeltema, Keith Bechtol, Aaron Roodman, and Pat Burchat. Each of these people has made a big impact on my experience in graduate school, and I feel so lucky to get to work with each of them. I am sure this list is incomplete and I apologize to anyone whose name I did not include!

I also need to thank the members of the Rubin Active Optics team. It is amazing what we have achieved together, and that we have had so much fun doing it! Sandrine Thomas has been an excellent and supportive leader; Josh Meyers is the most humble optics guru one could ask for; Chris Suberlak and Bryce Kalmbach have been wonderful colleagues in the trenches of the AOS pipelines; and Guillem Megias Homar has been irreplaceable as our summit expert. I am also lucky to count Guillem as one of my closest friends. Since meeting during an AOS telecon in December 2021, we have road tripped across Paraguay, gone boating off Costa Brava in Spain, toasted champagne as the sun rose over Cerro Pachón in Chile, and shared a large number of not-so-subtle laughs while texting during telecons. I look forward to a lifetime of friendship and scientific collaboration with him.

Last but not least, I would like to thank my friends and family for their love and support. I want to give a special shout out to my best friend and graduate school roommate, Rechel Geiger. We met while undergraduates at Duke University and moved to Seattle to start graduate school at the same time. We have been on this adventure together since day one, sharing the many ups and downs of graduate school and personal life. I am so proud of what

she has achieved and know she is equally proud of me.

I also want to thank my fiance, Michael Pearce, who I met while he was working towards his PhD in statistics at UW. Michael is a huge part of why the last few years have been the best of my life, and I am eternally grateful for his patience in planning a wedding with a partner who is writing their thesis and regularly disappearing to remote Chilean mountaintops.

DEDICATION

To my father, whose curiosity
and passion set me on this path.
I love and miss him dearly.

Chapter 1

INTRODUCTION

1.1 Tensions in the Era of Precision Cosmology

Over the last few decades a standard model of cosmology has emerged that, using only a few parameters, has had remarkable success in explaining a wide variety of astronomical observables over 14 billion years of cosmic evolution, including microscopic scales in the early universe to Gpc scales at late times (Fig. 1.1). This model, named Lambda-Cold Dark Matter (ΛCDM), has been tested by a wide range of cosmological probes, including, but not limited to, the primary and secondary anisotropies of the Cosmic Microwave Background (CMB) [232], chemical abundances produced by Big Bang Nucleosynthesis (BBN) [302], the ages of globular clusters [303], the evolution of the Baryon Acoustic Oscillation (BAO) scale [85], the redshift vs luminosity distance of Type Ia supernovae [36, 79], the clustering and weak lensing of galaxies [1, 209, 315], the growth and distribution of galaxy clusters and cosmic voids [80], galaxy rotation curves [278], the virial motion of cluster galaxies [183], redshift-space distortions [84], and the Lyman-alpha forest [61]. Most ΛCDM parameters are measured with near-percent level precision, a massive victory for the relatively young field of physical cosmology [228].

Despite these successes, ΛCDM remains phenomenological at heart, as the fundamental nature of two key ingredients remains a mystery. Cold dark matter, a non-relativistic energy density that clusters under gravity, interacts, at most, very weakly with light, and comprises approximately 80% of the matter in the universe, exerting a profound gravitational force on the baryons that form stars and galaxies. The cosmological evidence for dark matter is diverse and compelling. Originally hypothesized to explain the orbital velocity of stars in spiral galaxies [247, 248] and the virial motion of galaxies in the Coma cluster [325], today

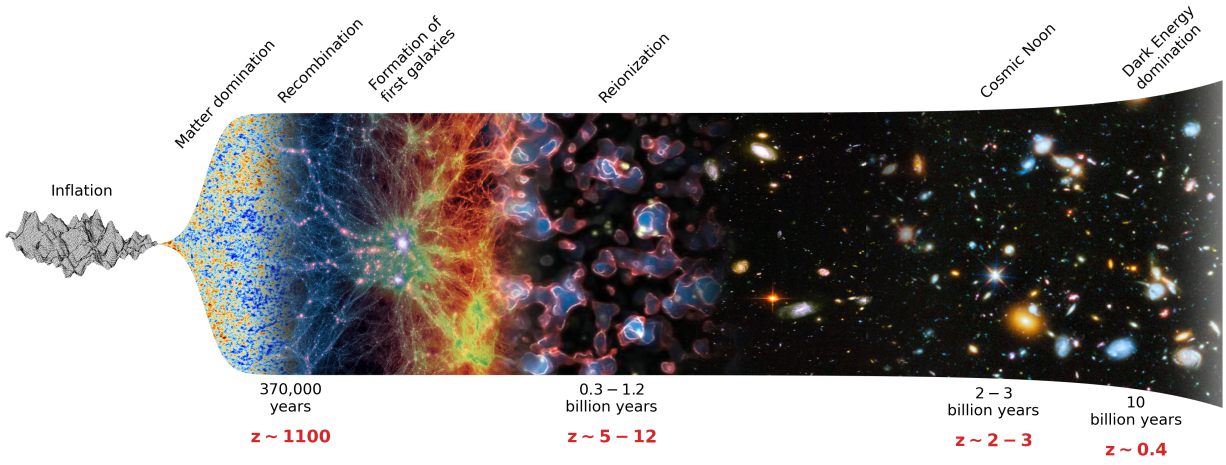

Figure 1.1: A schematic history of the universe, starting with exponential expansion during the epoch of inflation, from which quantum fluctuations seed macroscopic density perturbations in the universe. Inflation ends with reheating, starting the Hot Big Bang, during which the universe was filled with a hot, dense primordial plasma. The universe continues to expand at an ever decreasing rate, and the primordial plasma cools to the point that photons and baryons decouple, generating the cosmic microwave background (CMB) in an event known as recombination, after which almost all matter in the universe is electrically neutral. The CMB redshifts outside the visible range, and the universe enters a dark age. Density perturbations grow under gravity, forming the first stars and galaxies. UV radiation emitted by newly formed stars reionize the universe. Galaxies continue to form at an increasing rate until the epoch of cosmic noon, after which structure formation slows. At late times, dark energy begins to dominate the energy density, and the universe enters another era of accelerating expansion. This diagram is a composite of images from [12, 19, 231, 305].

it is recognized dark matter is a key ingredient in galaxy formation, and its presence is most clearly revealed by the CMB power spectrum and in gravitational lensing [56]. There are a number of promising theoretical candidates for the microphysical nature of dark matter, ranging ~ 70 orders of magnitude in mass, from axion and wave-like dark matter, to weakly interacting massive particles (WIMPs), to primordial black holes [18, 26, 64, 150]. Many years of searching in particle experiments have carved away large areas of that parameter space, however the true nature of the CDM remains unknown.

The Λ of ΛCDM refers to the cosmological constant – the ubiquitous, time-independent vacuum energy hypothesized by the Standard Model of particle physics to fill all of space. This component drives the expansion rate of the universe to accelerate at late times ($z \lesssim 0.7$). Evidence for this component began to grow throughout the 1980's and 90's as the measured matter density of the universe was found to be far insufficient to explain the age of the universe required by the oldest known stars [27] and the growth of large scale structure [97, 164], and the measured flatness of the universe demanded the total energy density exceed the measured matter density by a factor of ~ 5 [285]. Measurements of the redshift and luminosity distance to type Ia supernovae published in 1998 [229, 239] provided conclusive evidence the expansion rate of the universe was increasing. The vacuum energy hypothesis, however, faces huge challenges as the natural density of this energy component predicted by the Standard Model of particle physics exceeds the measured value by 120 orders of magnitude [28, 40], and indeed is many orders of magnitude too large to accommodate life as we know it [258, 309]. Understanding the nature of dark energy is seen by many to be the biggest challenge in physical cosmology, and indeed the largest observational cosmology experiments of the last two decades have been organized around this search [10].

As cosmological measurements get more precise, however, tensions have emerged, perhaps providing clues to a more fundamental theory underlying the phenomenology of ΛCDM. In particular, while ΛCDM is able to explain data from any individual probe, the combination of data from multiple probes has begun to strain the model, as each probe pushes the model to different regions of parameter space. The most famous of these tensions, the

Hubble (or H_0) tension, refers to a disagreement in measurements of the current expansion rate of the universe when comparing direct measurements in the local universe to indirect measurements of the early universe [89]. The direct method constrains H_0 by comparing redshifts of galaxies to their distances measured using the distance ladder, a system that relies on parallax and stellar astrophysics, while the indirect method derives H_0 using CMB anisotropies to constrain the distance to the surface of last scattering. This latter method requires one to assume a cosmological model, but measurements of CMB anisotropies are considered pristine and well-understood compared to the relatively messy world of variable stars, giant star evolution, and supernova explosions. Well-studied and viscously debated for over a decade, this tension has risen to $\sim 5\sigma$ significance and currently shows little sign of resolution, as new methods and new data from JWST show strong agreement with existing measurements [240], and proposals for new physics, such as early dark energy, face an ever-shrinking region of viable parameter space[1] [133, 143].

Another notable tension from recent years is the S_8 tension, which refers to a disagreement in the amplitude of (cosmologically) small-scale structure at late times. Constraining S_8 using weak lensing (measured via the shape distortions of distant galaxies) yields a smaller value of this amplitude compared to values inferred from CMB anisotropies (again assuming a cosmological model). This tension has never risen above $\sim 3\sigma$ significance, but has driven interesting inquiries into mechanisms for suppressing the amplitude of non-linear structures using stronger-than-expected baryonic feedback or new dark matter physics [e.g., 13]. Recent results from the Kilo-Degree Survey (KiDS), however, indicate this tension may be due to systematic errors in the methods used to measure the distance to galaxies in photometric surveys (i.e., photometric redshifts, or photo-z's) [315, 316]. While more investigation is required before declaring the S_8 tension dead, this demonstrates the care that must be taken in interpreting data from cosmological surveys to separate systematic errors from signs of new physics.

[1] Although recent results may indicate early dark energy is still viable, and maybe even preferred by the data [236].

1.2 A New Generation of Tensions

Within the last year and a half, a suite of new tensions has emerged, further straining the ΛCDM model. These tensions are driven, in large part, by measurements of the BAO scale made by the Dark Energy Spectroscopic Instrument (DESI; [175]), supported by supernova measurements from a variety of surveys [36, 79, 246] and photometric BAO measurements made by the Dark Energy Survey (DES; [81, 289]). These tensions go beyond the disagreement between early- and late-time measurements of ΛCDM parameters by questioning the fundamental assumptions of ΛCDM.

1.2.1 Discordant matter density

The first of these tensions is a modest disagreement in the measured fractional matter density of the universe, Ω_m. DESI BAO favor $\Omega_m = 0.295 \pm 0.015$, which narrows to $\Omega_m = 0.2962 \pm 0.0095$ when including full-shape measurements of the matter power spectrum [98]. This is in contrast to higher values favored by recent SNe experiments, which report $\Omega_m = 0.353 \pm 0.017$ (DES-SNY5; [79]), $\Omega_m = 0.331 \pm 0.018$ (Pantheon+, [36]), $\Omega_m = 0.359 \pm 0.027$ (Union3, [246]), representing a tension ranging from $1.5-2.9\sigma$ with DESI measurements. As this disagreement does not quite rise to the 3σ level, and Planck CMB constraints ($\Omega_m = 0.315 \pm 0.007$) lie between the two, it would be fair to argue that naming this a tension is over-eager. This Ω_m disagreement, however, hints at inconsistencies between these datasets, and, as we will see below, is deeply connected to more concerning tensions raised by these datasets.

1.2.2 Tensions in neutrino mass

Constraints on the sum of neutrino masses, Σm_ν, represent the second major tension in these datasets. Neutrinos are the lightest of the known fermions and only upper limits exist on their mass. Among lab experiments, KATRIN provides the tightest limit of $< 0.45\,\mathrm{eV}$ [157]. We know, however, at least two of three neutrino species must posses non-zero mass as their mass splitting gives rise to the neutrino oscillation phenomenon observed in terrestrial

experiments [6–8, 16, 110]. Neutrino mass also manifests in two separate ways in cosmology: by altering the geometry (i.e., expansion history) of the universe, and by suppressing the late-time growth of structure. A huge number of relativistic neutrinos are produced at early times, which then cool as the universe expands, becoming non-relativistic some time after recombination. Neutrinos, as a result, contribute to the radiation density at early times but contribute to matter density at late times. This late-time boost in the matter density delays the transition to dark energy domination, altering the distance vs redshift relation at low redshifts. In addition, while becoming non-relativistic at late times, neutrinos remain "hot" in that they have Mpc-scale free-streaming lengths, suppressing the growth of structure below this scale.

DESI constrains the neutrino mass primarily via the geometric effect[2]: the shape of the distance-redshift curve constrains Ω_m at late times, from which Σm_ν is inferred by subtracting the baryon and CDM masses constrained at early times by the CMB [98, 187]. DESI's constraint, $\Sigma m_\nu = -0.19^{+0.11}_{-0.16}\,\mathrm{eV}$, prefers a negative neutrino mass (equivalent to preferring *less* mass at late times than in the early universe), setting a 95% upper limit $\Sigma m_\nu < 0.053\,\mathrm{eV}$, lower than the 59 meV minimal neutrino mass allowed by neutrino oscillation experiments. DESI constraints on neutrino mass, therefore, are in tension with the fundamental assumptions of ΛCDM and with constraints from terrestrial particle physics experiments, driven in large part by DESI's preference for a lower Ω_m than the value preferred by the CMB. It remains to be seen whether these tensions are the result of some unknown systematic error or genuine signs of new physics, e.g., evolving dark energy [98, 234], modified gravity [20], time-varying neutrino mass due to cosmological phase transitions [185], new neutrino interactions [17, 102] (including with dark matter [122, 275]), additional long-range forces [67], or decaying dark matter [235].

[2]DESI also constrains neutrino mass via the suppression of structure, but the geometric constraint is dominant [98, 187].

1.2.3 Evolving dark energy

Evidence for the evolution of dark energy represents the third major tension in these data sets. Dark energy accelerates the expansion of the universe, altering the distance-redshift relation and damping the growth of structure at late times. The phenomenology of dark energy is typically described in terms of the equation of state that parametrizes the ratio of pressure to energy density: $w \equiv P/\rho$. The energy density of dark energy, normalized to its present value, is then

$$f_{\rm DE}(z) \equiv \frac{\rho_{\rm DE}(z)}{\rho_{\rm DE,0}} = \exp\left[3\int_0^z [1+w(z')]\frac{dz'}{1+z'}\right]. \tag{1.1}$$

A field with $w = -1$ corresponds to a cosmological constant, whose energy density is constant with time; for values $w > -1$, the density of dark energy decays with time, while for values $w < -1$, the so-called phantom regime, the density of dark energy grows with time.

Constraints on the evolution of dark energy are typically presented in the Chevallier-Polarski-Linder (CPL) parameterization [51, 177], a first-order Taylor expansion in the scale factor a:

$$w = w_0 + w_a(1-a), \tag{1.2}$$

where w_0 is the value of w at $z = 0$, w_a parametrizes the evolution of dark energy, and the scale factor $a = 1/(1+z)$. The normalized energy density in this model can be written

$$f_{\rm DE}(z) = (1+z)^{3(1+w_0+w_a)} \exp\left(-3\,w_a \frac{z}{1+z}\right). \tag{1.3}$$

In the CPL parameterization $w_0 = -1, w_a = 0$ corresponds to a cosmological constant. Note that while connecting constraints on this parameterization to specific models for evolving dark energy can be problematic [314], this parameterization still provides a useful diagnostic for detecting evolution of dark energy as a cosmological constant should still be consistent with $w_0 = -1, w_a = 0$.

Recent BAO and SNe Ia measurements suggest dark energy is evolving in time, at approximately $> 3\sigma$ tension with ΛCDM under the CPL parameterization [62, 81]. This conclusion

has proven robust in the face of more data collection (DESI year 1 vs year 2), sub-sampling data sets (removing DESI data most in tension with ΛCDM), swapping data from different surveys (DESI vs SDSS) and different survey types (spectroscopic vs photometric BAO), restricting to certain probes (excluding SNe or CMB data), and fitting different parametric and non-parametric models for the evolution of dark energy. In addition, these results indicate above redshift $z \sim 0.5$ dark energy is in the phantom ($w < -1$) regime, i.e., its energy density is *growing* with time, with several studies concluding this behavior is robust to the parameterization of $w(z)$ [62, 179]. This phantom behavior poses a number of theoretical problems, including vacuum instabilities, superluminal modes, Lorentz violation, and non-locality [192]. It is well known, however, that more complex models can appear as an effective fluid with $w < -1$ without actually requiring a phantom field, such as multifield models [138], models with late-time vacuum phase transitions [42, 88, 224] models with interactions between dark energy and dark matter [47, 76, 114], theories with modified gravity [44, 53, 103, 318], and theories that combine a negative cosmological constant[3] with a standard quintessence field [5, 200–202]. Regardless of the ultimate explanation, if confirmed, evolving dark energy would be the most important discovery of the last 25 years of cosmology.

The potential for evolving dark energy is also interesting in that it significantly increases the freedom afforded to the cosmological model, simultaneously enabling new solutions to other cosmological tensions. X. T. Tang et al. [288] find that evolving dark energy has the potential to solve the Ω_m tension, and that, in fact, fitting ΛCDM to a universe that evolves according to DESI's best-fit model naturally leads to Ω_m as constrained by SNe (BAO) to be biased high (low). Additionally, allowing for the evolution of dark energy significantly relaxes tension with neutrino mass, allowing DESI + CMB to prefer positive neutrino masses, and pushing the 95% upper limit to $\Sigma m_\nu < 0.177\,\mathrm{eV}$, easily accommodating the lower limits

[3] Note these models also purport to explain the unexpectedly large abundance of high-redshift galaxies observed by JWST [33, 106, 198], while simultaneously satisfying the "no de Sitter vacua" conjecture of string theory [75].

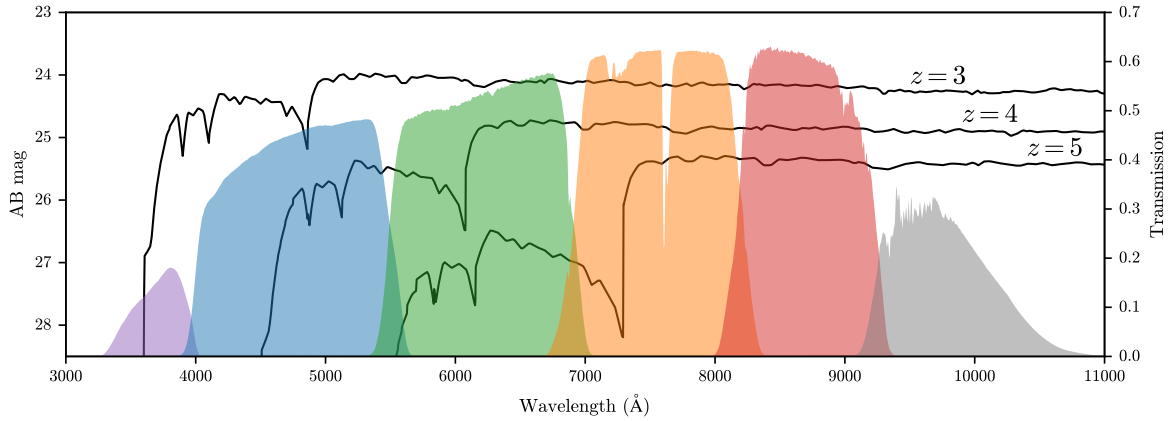

Figure 1.2: An example LBG spectrum at redshifts z =3, 4, 5, plotted with the Rubin filter bandpasses (left to right: u, g, r, i, z, y). As the LBG moves to higher redshifts the IGM absorption at restframe $\lambda < 1216\,\text{Å}$ shifts into longer wavelength bands. At wavelengths longer than this, the LBG spectrum is flat, reflective of ongoing star formation.

imposed by neutrino oscillation experiments [98]. Many of these models for evolving dark energy also have the potential to solve the older Hubble and S_8 tensions [88, 103, 114].

The interrelation of these tensions may be an intriguing clue that they are real and driven by some new physics, but it is also a warning that several of these tensions may be driven by the same underlying systematic error. Regardless, the emergence of these new tensions provides new excitement and urgency for the major new cosmology surveys arriving this decade.

1.3 Probing the High-redshift Universe with Lyman-break Galaxies

Previous generations of precision cosmology experiments, including those driving cosmological tensions, have primarily probed the $z < 2$ and $z \geq 1100$ universe, while the epoch in between remains largely unconstrained. The highest-redshift constraints that currently push into this regime are spectroscopic observations of quasars and the Lyman-alpha forest, which nominally reach redshifts of $z \lesssim 3$, but only with number densities of order $10\,\text{deg}^{-2}$ [61].

This intervening epoch represents a huge volume in which structure remains linear to much smaller scales, providing great potential for constraining cosmology [312], and in particular provides unique leverage for constraining several key quantities that are subject to the tension described in the previous section.

The $2 < z < 6$ universe is especially interesting as it can be probed by upcoming galaxy surveys[4] via the selection of Lyman-break galaxies (LBGs). LBGs are high-redshift star-forming galaxies whose rest-frame UV flux below $\sim 1216\,\text{Å}$ is absorbed by neutral hydrogen in stellar atmospheres, the interstellar medium (ISM), and the intergalactic medium (IGM). This absorption creates a distinctive break in the galaxy spectrum (i.e., the Lyman break) which is easily identifiable in photometric surveys (Figure 1.2). This technique, pioneered in the 1990's [282], has been the primary method with which to study galaxies the high-redshift universe for the last 30 years (e.g., [30, 31, 104, 105, 223]).

In recent years this method has been applied to data from JWST to study the highest-redshift galaxies ever discovered [45, 92, 93, 106, 130, 215]. These galaxies, at $z > 10$, are some of the first galaxies to form in the universe, in an epoch called Cosmic Dawn. Their unexpectedly large number densities have challenged our previous models of galaxy formation [33, 106, 198] and continue to drive a revolution in our understanding of galaxy formation in the first billion years of cosmic history.

Previous LBG surveys have primarily been deep and very narrow, focused on the astrophysics of LBGs, but current and upcoming cosmology surveys are beginning to provide the width and depth to deliver high number densities of LBGs across huge areas of the sky. The GOLDRUSH program of the Hyper Suprime-Cam (HSC) Subaru Strategic Program (SSP) has been a pioneer in this respect, discovering $\sim 4 \times 10^6$ LBGs across $300\,\text{deg}^2$ of the sky, probing the star formation history, galaxy-halo connection, and the amplitude of structure at high redshifts [128, 129, 206, 223, 300]. The Vera C. Rubin Observatory's Legacy Survey

[4]Line-intensity mapping, including 21 cm measurements, promises to push to much higher redshifts, but these probes have not yet reached the level of maturity where they can produce constraints on cosmological parameters.

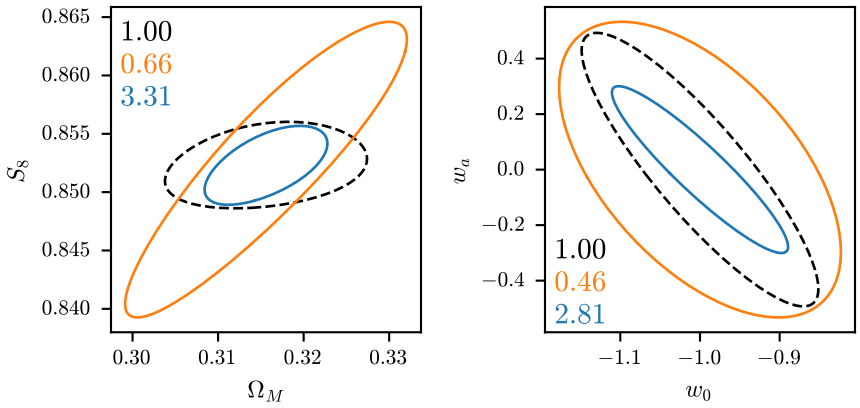

Figure 1.3: Constraints on cosmology using high (gold) and low (black) redshift 3×2pt analyses. The blue contours show the joint constraints. The numbers in each panel represent the figure of merit (i.e., inverse area of the the ellipse) relative to the low-redshift constraint. The left panel forecasts S_8 vs Ω_m in year 1 of LSST; the right panel forecasts w_a vs w_0 in year 10 of LSST. In both cases the high-redshift constraints are weaker in isolation, but yield significant improvements when combined with the low-redshift constraints. See Chapter 5 for more details.

of Space and Time (LSST; see the next section) is expected to detect LBGs at a number density of order 10,000 $\deg^{-2}$ in the range $2 < z < 6$ across 18,000 $\deg^2$ of sky (See Chapter 4), increasing the number density of high-redshift tracers by a factor of 10^3.

The clustering of these LBG populations, and their cross-correlation with the weak gravitational lensing of the CMB, provide an extremely powerful probe with which to constrain cosmology, especially when jointly analyzed with low-redshift constraints. Figure 1.3 shows forecasts for constraints on cosmological parameters using high (gold) and low (black) redshift data from LSST. In both cases the high-redshift constraints are weaker in isolation, but yield significant improvements when combined with low-redshift constraints in a joint analysis (blue). In the left panel one can see the constraints on Ω_m become significantly tighter due to the different degeneracy directions exhibited by the high- and low-redshift

constraints in the S_8 vs Ω_m plane.

The right panel shows that the degeneracy directions of the high- and low-redshift constraints are very similar in the w_a vs w_0 plane, but the addition of high-redshift data nevertheless improves the joint constraint to a greater extent than might be expected by looking at the size of the contours. This is because when constraining the long-term evolution of dark energy adding data at high-redshifts provides significant leverage by extending the redshift baseline of the constraints. These constraints on the evolution of dark energy are driven by constraints on the growth of large scale structure between high and low redshifts. This is a complementary method to the geometric constraints provided by BAO and SN surveys, and therefore provides a useful cross-check of systematics in both methods. These forecasts demonstrate that high-redshift data from LSST will be very powerful for determining whether dark energy is truly evolving, whether it has phantom behavior at high redshifts, and how that phantom phase evolves, which will help determine which theory for evolving dark energy is correct [178, 179].

High-redshift LBGs also provide a unique opportunity to constrain the sum of neutrino masses via their redshift-dependent suppression of structure growth. These measurements, which are complementary to the geometry-driven constraints, require a high-redshift constraint on large scale structure with which to compare the amplitude of low-redshift structure. The high-redshift constraint typically relies on the CMB, however the amplitude of the CMB power spectrum is degenerate with the optical depth of reionization (τ), the least well constrained CMB parameter. LBGs, however, provide a means with which to constrain high-redshift structure that is independent of τ, enabling neutrino mass constraints with precision 25 meV, independent of the CMB, BAO, or SN measurements [320] and competitive with the cutting-edge DESI constraints [98]. This is especially promising as systematics associated with τ could be responsible for the neutrino mass and evolving dark energy tensions[5] [257]. LBG constraints on high-redshift structure, therefore, will provide a powerful cross-check to

[5]Although relieving these tensions via modifying τ would create tensions with measurements of Lyman-alpha constraints in the IGM [214, 216].

determine the effective sum of neutrino masses constrained by cosmology is truly below the lower limit on neutrino mass set by oscillation experiments, as well as whether systematics associated with τ could be driving multiple cosmological tensions.

Additionally, as noted multiple times in this and the previous section, the Ω_m, neutrino mass, and evolving dark energy tensions are all interconnected. These new constraints provided by high-redshift LBGs, therefore, will have multiple cross-cutting impacts, helping to untangle the connections between these tensions.

Beyond the tensions discussed here, other authors have forecast the power of LBGs to constrain the high-redshift evolution of S_8 at the 1% level [312], two orders of magnitude better than existing constraints [206], providing a model-agnostic probe of modified gravity [2, 212]. Via measurement of scale-dependent galaxy bias, LBG clustering can also constrain the amplitude of local-type primordial non-Gaussianity ($f_{\rm NL}$) with $\sigma(f_{\rm NL}) \sim 1$ precision [269]. This is 5 times better than current best constraints [233], and reaches the typical benchmark of $f_{\rm NL} \sim 1$, providing a sensitive probe of inflation, and in particular can determine whether inflation is driven by multiple quantum fields [46].

LBGs detected by LSST and other upcoming surveys will also dramatically improve our understanding of galaxy formation and evolution in the high-redshift universe. The epoch accessible by LSST, the $2 < z < 7$ universe, is especially interesting as it connects the end of reionization at $z \sim 6$ to the peak of star formation at $z \sim 2$ (en epoch known as cosmic noon; Figure 1.1). This era sees a rapid rise in the rate of star formation, in the abundance of massive, luminous galaxies, and in the number of active galactic nuclei (AGN). Constraining number densities and the luminosity function of LBGs across this epoch will enable precise constraints on the history of (unobscured) star formation [129], the galaxy halo connection [128, 142], and the onset of AGN feedback [32, 104, 223]. The luminosity function at $5 \lesssim z \lesssim 7$ will also provide insight into the galaxy populations that drive reionization, informing whether reionization is a slow, smooth process, or whether it is a late, quick, and patchy process [121, 167, 214, 216, 277]. The bright end of the luminosity function can also provide constraints on $f_{\rm NL}$ that complement constraints from scale-dependent galaxy bias

[254–256].

This is a very exciting time to be studying the high-redshift universe. JWST has now been continuously revolutionizing our understanding of high-redshift galaxies for 4 years. Rubin's LSST will provide wide and deep photometry across 6 optical bands, enabling the detection of hundreds-of-millions of LBGs. The Euclid and Roman Space Telescopes [263, 279] will provide deep infrared photometry, improving the purity and redshift calibration of these samples. Later this decade, DESI-II is expected to collect spectra for ∼40 million LBGs. Optimizing LBG cosmology will require combining all of this data.

1.4 The Rubin Observatory's Legacy Survey of Space and Time

The Vera C. Rubin Observatory sits on Cerro Pachón in the Coquimbo region of the Chilean Andes ([146, 189]; Figure 1.4). Using the Simonyi Survey Telescope, equipped with an 8.4 m primary mirror and a 3.2 gigapixel camera (Figure 1.5), Rubin will conduct the Legacy Survey of Space and Time (LSST), the largest astronomical survey ever undertaken. Over the course of 10 years, LSST will image roughly 18 000 square degrees of the southern sky to an unprecedented depth[6] for such a wide area (Figure 1.6). This survey will detect billions of galaxies, including approximately 100 million at $z > 2$, enabling cutting-edge constraints on cosmology and astrophysics [189].

Rubin's Simonyi Survey Telescope is a three-mirror anastigmat, designed to deliver high optical quality across a 9.6 deg^2 field of view. Rubin images in 6 bands (*ugrizy*; Fig. 1.2) ranging from the near UV to the near IR. The cutoff on the UV side is due to Rayleigh scattering in the atmosphere and the ultraviolet cutoff in the reflectivity of silver (which coats the surface of all three mirrors). The cutoff on the IR side is due to the silicon CCDs becoming transparent to long-wavelength photons, as these do not posses sufficient energy to overcome the band gap of the silicon. These 6 bands provide strong leverage for photometric redshift (photo-z) estimation, i.e. determining the distances to galaxies using

[6]Recent forecasts of the LSST Wide Fast Deep survey suggest 10 year 5σ point source depths of 25.36, 26.60, 26.79, 26.32, 25.66, 24.73 in the *ugrizy* bands, respectively [72].

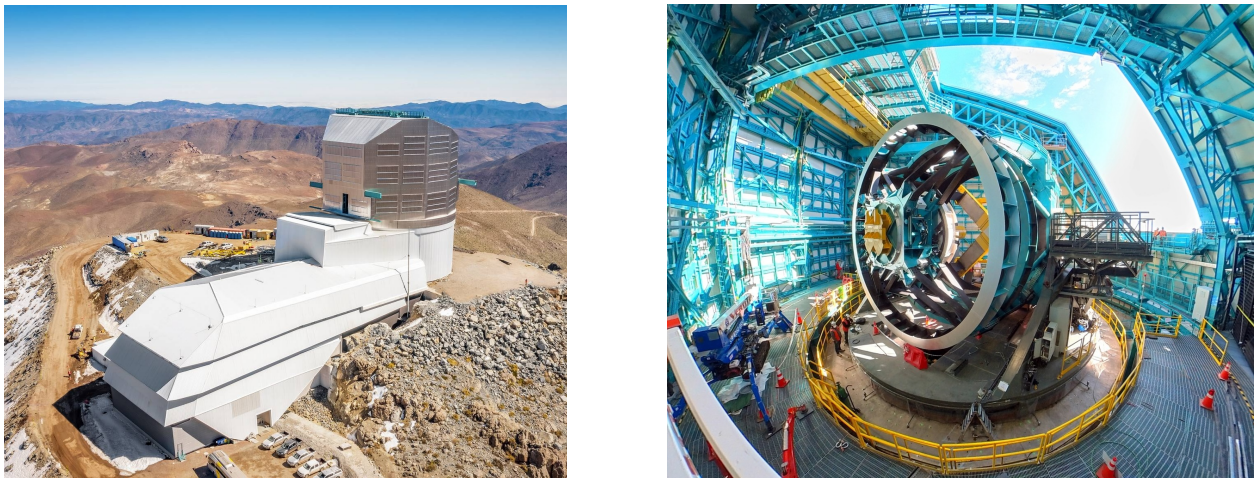

Figure 1.4: Left: The Vera C. Rubin Observatory on Cerro Pachón in the Chilean Andes. Right: The Simonyi Survey Telescope inside the Rubin Observatory dome. Image credits: A. Pizarro D./RubinObs/NOIRLab/SLAC/NSF/DOE/AURA; H. Stockebrand/RubinObs/NSF/AURA.

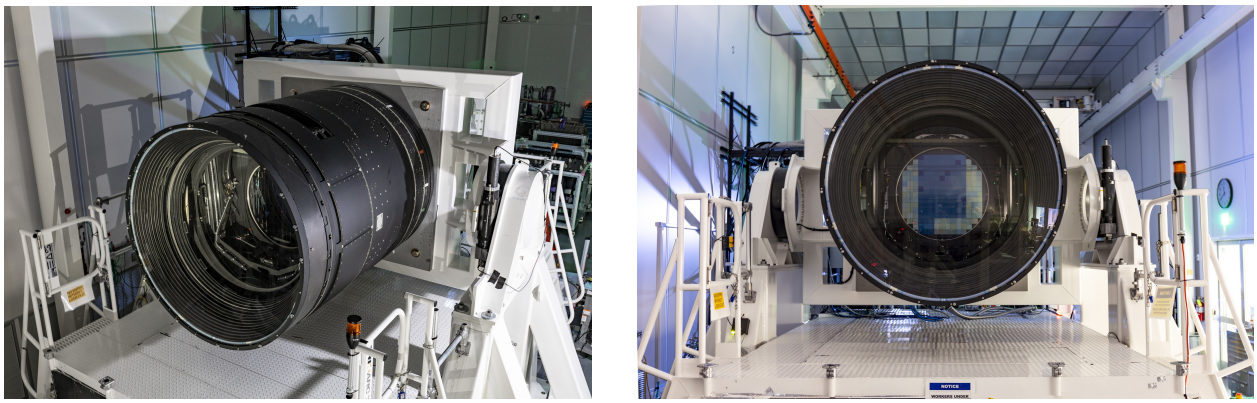

Figure 1.5: LSSTCam, the camera which will be used to conduct LSST. On the right you can see the focal plane of the camera, reflected in blue. The two different colors correspond to CCDs supplied by two different manufacturers. Image credit: Jacqueline Ramseyer Orrell/SLAC National Accelerator Laboratory.

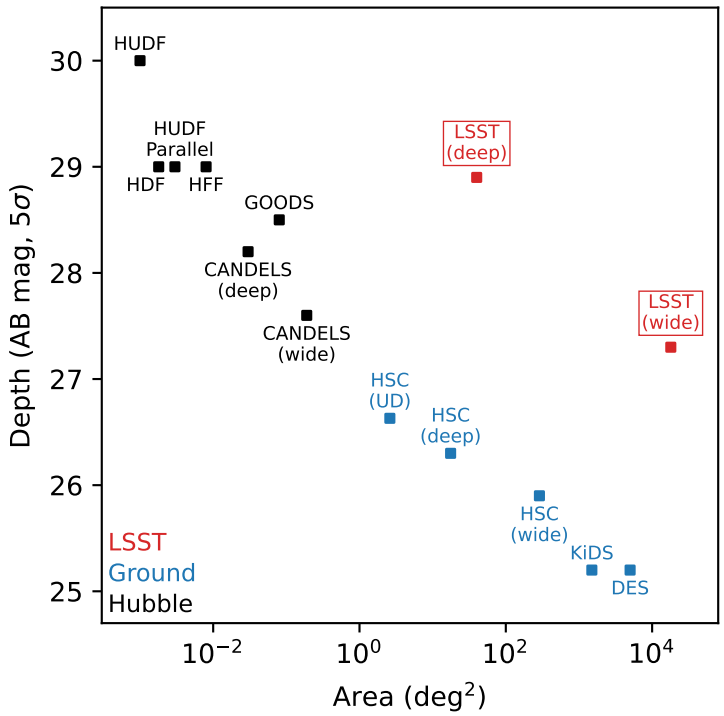

Figure 1.6: The width and depth of LSST compared to precursor surveys from space (black) and from the ground (blue). These surveys all lie on a line representing the trade-off between survey width and depth. LSST represents a huge step off this line into a regime of unprecedented width *and* depth.

only photometric colors [156, 189]. In particular, Rubin's 6 bands provide strong localization of the Balmer break (at $\lambda \sim 4000$ Å) for redshifts $z \lesssim 1.6$ and the Lyman break (at $\lambda \sim 912$ Å) for redshifts $2.5 \lesssim z \lesssim 8$.

Rubin began on-sky commissioning in late 2024, during which time I spent 14 nights on the summit aiding in the commissioning of the telescope and in data collection for the science validation survey using the Commissioning Camera (ComCam).

The Rubin Observatory employs an active optics system (AOS) to maintain optical alignment and mirror figure required to deliver pristine optical quality across its wide field of view.

A large fraction of my research time during my PhD has been dedicated to this system. See the appendices for more information on this work.

1.5 Structure of the Thesis

This thesis consists of my PhD work dedicated to preparing for precision cosmology with LSST, with a focus towards enabling high-redshift constraints using LBGs. Chapters 2 and 3 detail work on improving systematic errors associated with photo-z estimation, one of the most important sources of uncertainty in photometric galaxy surveys. The former develops an algorithm for learning representative galaxy spectra directly from broadband photometry, and demonstrating that these spectral templates improve photo-z estimation. The latter develops a deep learning algorithm for forward modeling galaxy populations in a probabilistic way, enabling more detailed and consistent studies of errors associated with photo-z estimation. Chapter 4 focuses on optimizing the LSST survey strategy for photo-z estimation and the detection of LBGs. This chapter also makes forecasts for the number of LBGs that will be detected by LSST. Chapter 5 forecasts cosmology constraints using LBG clustering and their cross-correlation with CMB lensing. This chapter explores important sources of systematic error, and makes recommendations for early science with LSST. I conclude in Chapter 6.

The main body of this thesis is focused on photo-z estimation, LBG detection, and forecasts for LSST cosmology. However a significant fraction of my time during my PhD has been dedicated to development and commissioning of the Rubin Observatory active optics system. For information on this work, see the appendices.

Chapter 2

LEARNING SPECTRAL TEMPLATES FOR PHOTOMETRIC REDSHIFT ESTIMATION FROM BROADBAND PHOTOMETRY

Photometric redshift (photo-z) estimation is a vital component to nearly all cosmology probes in photometric surveys. There are two main schools of photo-z estimators: machine learning and template-based estimators. The former provides the most precise and accurate results when trained on a spectroscopic sample that is representative of the colors and redshifts present in the photometric target sample [218]. Representative spectroscopic samples, however, are frequently not available, especially when pushing to great depths as LSST will do. This is also true for high-redshift galaxy samples, for which spectroscopic data is severely limited. As a result, cosmology with high-redshift LBGs will need to rely on template-based methods.

Template-based photo-z estimators rely on a physical model for galaxy spectra, and compare observed photometry to synthetic photometry calculated for spectral templates produced by the model. These models are less precise and accurate than machine learning methods when applied to targets within the remit of spectroscopic training sets, but have better performance and smaller biases when extrapolating beyond the color and redshift range of the spectroscopic data. There are important and somewhat ambiguous modeling choices, however, that must be made when assembling the template set or setting the priors for the galaxy SED model. In addition, observed galaxy spectra are often used in the construction of these template sets, but the spectro-photometric calibration of these spectra is often difficult. In other words, it is difficult to mutually calibrate spectroscopic and photometric observations to the precision required for accurate photo-z estimation.

The following chapter, originally published in *The Astronomical Journal 160, 191 (2020)* [69], seeks to bridge the gap between machine learning and template-based methods to remove these ambiguities and inadequacies by learning spectral templates for photo-z estimation directly from broad band photometry. We demonstrate this method is able to resolve high-frequency spectral structure far beyond the resolution of the input broadband photometry, and that photo-z's estimated with these templates are more accurate and less biased.

This method does rely on the use of spectroscopic redshifts (spec-z's). We believe, however, more advanced models can be built that do not rely on spec-z's, which will extend the utility of this method to higher redshifts. This possibility is discussed in Chapter 6.

2.1 Abstract

Estimating redshifts from broadband photometry is often limited by how accurately we can map the colors of galaxies to an underlying spectral template. Current techniques utilize spectrophotometric samples of galaxies or spectra derived from spectral synthesis models. Both of these approaches have their limitations, either the sample sizes are small and often not representative of the diversity of galaxy colors or the model colors can be biased (often as a function of wavelength) which introduces systematics in the derived redshifts. In this chapter we learn the underlying spectral energy distributions from an ensemble of ∼100K galaxies with measured redshifts and colors. We show that we are able to reconstruct emission and absorption lines at a significantly higher resolution than the broadband filters used to measure the photometry for a sample of 20 spectral templates. We find that our training algorithm reduces the fraction of outliers in the derived photometric redshifts by up to 28%, bias up to 91%, and scatter up to 25%, when compared to estimates using a standard set of spectral templates. We discuss the current limitations of this approach and its applicability for recovering the underlying properties of galaxies. Our derived templates and the code used to produce these results are publicly available in a dedicated Github repository: https://github.com/dirac-institute/photoz_template_learning.

2.2 Introduction

Studies of galaxy evolution, galaxy clusters, large-scale structure, weak lensing, etc all rely on the determination of galaxy redshift. Spectroscopic surveys of galaxies can provide very accurate redshifts by measuring the shifted wavelengths of sharp spectral features such as emission and absorption lines. Despite advancements in multi-object spectrographs, spectroscopic measurements are expensive and time-consuming and we can only collect spectra for a small fraction of the galaxies that can be imaged by modern surveys, such as the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 290) and the Kilo-Degree Survey (KiDS; J. T. de Jong et al. 78). This problem will only increase in magnitude as the next generation of surveys, such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 189) and the Wide-Field Infrared Survey Telescope (WFIRST; J. Green et al. 123), image orders of magnitude more galaxies at fainter magnitudes than are present in current data sets. As a result, rather than rely on spectroscopic redshifts (spec-z's), modern surveys increasingly rely on photometric redshifts (photo-z's; see M. Salvato et al. 260 for a review).

Photo-z's are estimates of galaxy redshifts derived from changes in the colors of galaxies as their spectral energy distributions (SED's) redshift through a series of broadband filters. This estimation is typically done using one of two approaches: machine learning (ML) or template fitting (see e.g. S. J. Schmidt et al. 268 for an evaluation of many examples of the two).

Machine learning approaches train on a data set of photometry with spec-z's and attempt to directly learn an empirical relationship between galaxy colors and redshift (e.g. A. J. Connolly et al. 63, `TPZ` M. C. Kind & R. J. Brunner 159, `FlexZBoost` R. Izbicki & A. B. Lee 148, `CMNN` M. L. Graham et al. 119). Once trained, they can predict galaxy redshifts given photometry alone. The advantage of ML methods is that the effects of dust, galaxy evolution, and other relevant variables are encoded in the training set and thus it is possible for ML methods to account for these in the derived mapping from colors to redshift if the data

encapsulate these effects. The success of this mapping depends on the choice and complexity of the ML model and the corresponding hyperparameters. The downside of ML methods is that their success relies on how representative and well-controlled the training set is, and that they are unable to extrapolate beyond that set.

Template fitting photo-z estimators (e.g. `LePhare` S. Arnouts et al. 15, `BPZ` N. Benítez 21, `EAZY` G. B. Brammer et al. 35) work on the assumption that galaxy photometry are sampled from a relatively small set of underlying spectral types, characterized by the eponymous SED templates. These estimators calculate photo-z's by selecting the template and redshift with simulated fluxes most similar to the observed fluxes. In order for this method to work, the underlying SED templates from which the galaxies are sampled must be known. Common methods for generating these templates include simulating galaxy SED's from spectral synthesis models, e.g. G. Bruzual A. & S. Charlot [38], and deriving templates from the observed spectra of local galaxies, e.g. N. Benitez et al. [22].

The primary advantage of template fitting methods is that it is not limited to the bounds of a training set. A key limitation is that they do not guarantee that the SED templates will span the full distribution of galaxy spectra in a given data set, nor that they will properly account for the effects of dust, or spectral evolution. In addition, spectral synthesis models are only able to generate spectra with a discrete set of physical parameters (e.g. temperature, age, metallicity), and obtaining real galaxy spectra is expensive, especially at the redshifts and magnitudes that will be observed by LSST.

Several previous works have attempted to combine the advantages of these two approaches by deriving SED templates from a photometric training set, and then using the derived templates for photo-z estimation (T. Budavári et al. 39, I. Csabai et al. 74). These approaches leverage a large set of galaxy photometry, which amount to low resolution spectra, to sample a smaller set of SED templates across a broad range of rest wavelengths. This effectively over-samples the template SED's, allowing us to reconstruct spectral features at a resolution much higher than that of the broadband filters used to measure the photometry. This is analogous to the Drizzle technique used to reconstruct higher resolution images for the

Hubble Space Telescope (HST; A. Fruchter & R. Hook 109) and the reconstruction of SED's using differential chromatic refraction (DCR; M. A. Lee et al. 172).

This template learning approach retains the physical motivation and extensibility of the template fitting method, while taking advantage of learning the systematics and confounding variables implicit in the training set. In addition, it opens the possibility of learning a smooth continuum of galaxy spectra, in contrast to the discrete set offered by the limited galaxy observations and galaxy modeling codes.

While previous works attempt to learn galaxy templates from data using a set of eigen-spectra, we adapt the algorithm of T. Budavári et al. [39] to directly learn a set of templates from the data.We extend these earlier works by applying our methods to a large data set of 102,476 galaxies with spec-z's and photometry in 19 bands. In this manner we are able to learn a variable number of SED templates with clear spectral features, and with simple postprocessing, we are able to further reconstruct emission lines in the bluest templates.

We show that templates can be learned from scratch or as perturbations of pre-existing templates. We use these learned templates to estimate photo-z's with `BPZ` and find that the training reduces the bias and scatter of the redshift estimates, with little impact on the fraction of catastrophic outliers. In addition, we find that both bias and scatter decrease with the number of SED templates used in the photo-z estimation.

The outline of the chapter is as follows: in Section 2.3 we describe the template training algorithm, including how to match photometry to templates, how to perturb templates to better match the photometry, and how to select the hyperparameters for training. In Section 2.4, we describe the spec-z and photometric data sets used in the template training and redshift estimation. In Section 2.5, we apply the template training algorithm to sets of naive templates and to a pre-existing set of templates derived from galaxy observations and spectral synthesis models. We discuss the performance of the algorithm including its convergence, and the accuracy of the reconstructions. In Section 2.6, we use our templates to estimate photo-z's for a training set of galaxies and analyze the results. We discuss our results and future goals in Section 2.7 and conclude in Section 2.8.

2.3 Template Training Algorithm

In this section, we will present an approach for learning SED templates directly from broadband photometry, using a modified version of the algorithm developed in T. Budavári et al. [39]. If we assume that the galaxies in our data set are sampled from a small set of underlying spectra, the SED templates, and we know the spectroscopic redshift for each galaxy, we can shift the photometry to the restframe and treat each observation of a redshifted galaxy as a *restframe observation* of one of the templates with a different set of effective filters. With a large enough data set, the wavelengths of the effective filters will overlap substantially. This over-sampling allows us to recover higher resolution features in the templates, even though the data are low resolution observations of different galaxies.

Let us assume we have a set of SED templates as a starting point, which can represent rudimentary guesses and need not resemble true galaxy spectra. In the first part of this section, we describe a method by which we create a training set of broadband photometry for each template from a large data set of galaxy photometry. In the second part, we derive the perturbation algorithm that is used to train each SED template on its corresponding photometry set. The full training algorithm is an expectation maximization that consists of iterating these two steps: matching photometry to templates, and perturbing templates to better match the photometry. This process is iterated until the SED templates converge. In the final part, we discuss a heuristic for selecting the training hyperparameters.

2.3.1 Matching Photometry Sets

Assume we have a set of naive SED templates and a large set of observed fluxes, $\{f_m\}$, with known spectroscopic redshifts, z_m. Our goal is to train each template on an appropriate subset of the $\{f_m\}$, so that the naive templates better represent the colors of the galaxies. To assemble these training sets, we consider subsets $\{f_n\} \subset \{f_m\}$, corresponding to the observed fluxes of a single galaxy at redshift z, where the subscript n denotes different

filters. We compare these observed fluxes with the template fluxes $\{\hat{f}_n\}$, where

$$\hat{f}_n = \int S\left(\frac{\lambda}{1+z}\right) R^n(\lambda) d\lambda, \tag{2.1}$$

$S(\lambda)$ is an SED template, and $R^n(\lambda)$ is the normalized response function of the filter used to measure the flux f_n. For photon counting detectors,

$$R(\lambda) = \frac{\lambda\, T(\lambda)}{\int \lambda\, T(\lambda) d\lambda}, \tag{2.2}$$

where $T(\lambda)$ is the system response function that captures the transmittance of the atmosphere and the response of the detector [24].

The observed fluxes are assigned to the template whose colors are most similar, which is determined by normalizing the observed and template fluxes in the same band and picking the template that minimizes the squared differences of the fluxes. The normalization band is chosen by selecting the band for which the ratio $\hat{f}_n/f_n$ is the median of the flux ratios for that galaxy. By performing this matching and renormalization for each galaxy in the photometry set, we associate a subset of the galaxies (and the corresponding photometry) to each template.

Examining how the galaxies are assigned to the individual templates is helpful in selecting the initial set of templates. The initial templates should be chosen so that the matching algorithm roughly divides the galaxies by their colors. It is also important that each set contains a sufficient number of fluxes distributed across the wavelengths of interest, as the perturbation algorithm derived in the next section relies on over-sampling to reconstruct higher resolution features of the SED templates.

2.3.2 The Perturbation Algorithm

Assume we have a set of photometry, $\{f_n\}$, which constitute observations of the same underlying SED template, $S(\lambda)$, at various known redshifts, z_n. These observed fluxes should approximately match the template fluxes calculated via Equation 2.1. However, we can also calculate the template fluxes by imagining that we are observing the template in its rest

frame using a set of effective, blueshifted filters:

$$\hat{f}_n = \int S(\lambda)\, R^n[(1+z_n)\lambda]\, d[(1+z_n)\lambda] \tag{2.3}$$

$$= \sum_k s_k\, r^n_{k'} \Delta\lambda_{k'}, \tag{2.4}$$

where in the second line s_k and r^n_k are the discrete representations of $S(\lambda)$ and $R^n(\lambda)$ respectively, parameterized by the wavelength bins $\{\lambda_k\}$ with widths $\{\Delta\lambda_k\}$. Primed indices indicate redshifted wavelengths, i.e. $\lambda_{k'} = (1+z_n)\lambda_k$ and $\Delta\lambda_{k'} = (1+z_n)\Delta\lambda_k$.

We wish to perturb the template so that the template fluxes, $\hat{f}_n$, better match the observed fluxes, f_n. Letting $\hat{s}_k$ be a new template resulting from a perturbation of s_k, we define the cost function (T. Budavári et al. 39 Equation 7):

$$\chi^2 = \sum_n \frac{1}{\sigma_n^2}(\hat{f}_n(\{\hat{s}_k\}) - f_n)^2 + \sum_k \frac{1}{\Delta_k^2}(\hat{s}_k - s_k)^2. \tag{2.5}$$

The optimum perturbation is found via a multidimensional minimization of the cost function. The first term in Equation 2.5 penalizes differences between the observed fluxes and the perturbed template fluxes, weighted according to σ_n (the fractional error of the measured flux). The second term in Equation 2.5 penalizes large perturbations, weighted by the hyperparameters Δ_k. This parameter controls learning rate and also helps stabilize the results. See the next section for more details.

We follow T. Budavári et al. [39] by introducing the simplifying perturbation and constant terms:

$$\begin{aligned} \xi_k &= \hat{s}_k - s_k \\ g_n &= f_n - \sum_k s_k\, r^n_{k'} \Delta\lambda_{k'}. \end{aligned} \tag{2.6}$$

Then, we have:

$$\chi^2 = \sum_n \frac{1}{\sigma_n^2}\left(g_n - \sum_k \xi_k\, r^n_{k'} \Delta\lambda_{k'}\right)^2 + \sum_k \frac{\xi_k^2}{\Delta_k^2}, \tag{2.7}$$

which can be analytically minimized:

$$\frac{\partial \chi^2}{\partial \xi_l} = 0 \implies \sum_k M_{lk} \tilde{\xi}_k = \nu_l. \tag{2.8}$$

The matrix M and vector ν are defined

$$\begin{aligned} M_{lk} &= \sum_n \frac{1}{\sigma_n^2} (r_{l'}^n \Delta\lambda_{l'})(r_{k'}^n \Delta\lambda_{k'}) + \frac{\delta_{lk}}{\Delta_k^2}, \\ \nu_l &= \sum_n \frac{g_n}{\sigma_n^2} (r_{l'}^n \Delta\lambda_{l'}), \end{aligned} \tag{2.9}$$

where δ_{lk} is the Kronecker delta. One can then solve for $\tilde{\xi}$. The perturbed spectrum is then $\hat{s}_k = s_k + \tilde{\xi}_k$.

Iterating the perturbation changes the shape of the template SED to better match the measured photometry, as shown in T. Budavári et al. [39]. An example of this process can be seen in Figure 2.1. Fluxes in the *ugrizY* filters listed in Table 2.2 were calculated for a starburst galaxy template at 1000 random redshifts $z < 3$. Starting with an $S(\lambda) = 0$ template SED, the perturbation algorithm is applied iteratively. After 100 iterations, the trained template closely matches the original template in the wavelength range for which photometry exists. While the trained template is a smoothed version of the original, high resolution features have been recovered, despite the relatively low resolution of the filters. In practice, higher Δ_k can be chosen so that fewer iterations are required in the training; a lower value was chosen here so that the effects of successive iterations can be more clearly seen. See Section 2.3.3 for further discussion of selecting the hyperparameters.

The perturbation algorithm changes the shape of the template SED's so that re-running the photometry matching will now result in different subsets of galaxies assigned to each template. The full training algorithm is iterated until the SED templates converge.

2.3.3 Selecting Hyperparameters

The success of the training algorithm depends on the chosen hyperparameters. The first is the number of templates. As discussed in Section 2.3.1, this choice can be made by using

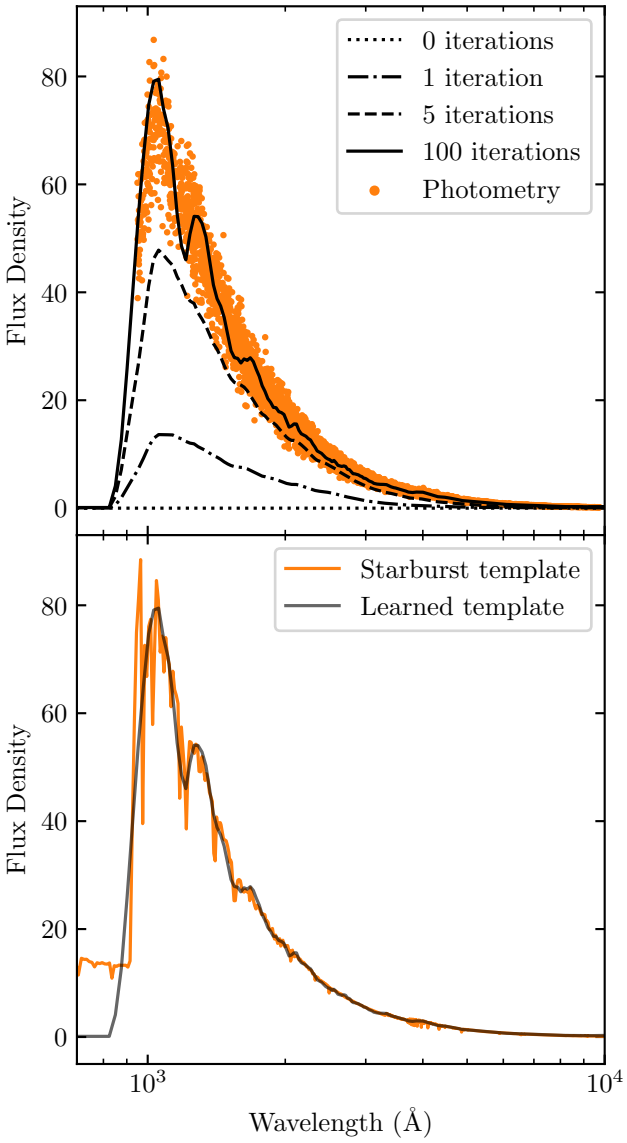

Figure 2.1: Perturbing a naive template, in this case a flat line, to better match a photometry set. Top: the orange points are simulated observations of the 5Myr starburst template from D. Coe et al. [59] at 1,000 random redshifts in the range z=0 to z=3 using the *ugrizY* filters listed in Table 2.2. The simulated photometry has 10% Gaussian error. The template is shown after various stages of the training. Bottom: the learned template is plotted with the original starburst template.

the photometry matching algorithm and choosing the appropriate number of templates to approximately separate out the different spectral shapes displayed in the photometry. For further discussion of how the number of templates affects photo-z results, see Section 2.6.3.

The rest of the hyperparameters consist of the set of Δ_k. These parameters, which set the relative weighting of the regularization term in Equation 2.5, determine the stability and speed of the training algorithm. If the Δ_k are too large, training will be very slow and a large number of iterations will be required. If the Δ_k are too low, the training becomes unstable and the final templates will be over-fit. Here we present a heuristic for selecting an appropriate value to balance these two extremes.

For the work presented below, the index k is dropped, so that $\Delta \equiv \Delta_k$ has a single value for each training set that is independent of wavelength. In choosing the appropriate value of Δ for each training set, it is desirable to select a value that corresponds to a constant ratio, w, of the flux and regularization terms in Equation 2.5. The necessary value of Δ will vary by training set, as the number of terms in the sum over fluxes (i.e. the sum over n in Equation 2.5) will vary by training set. To this end, we make the following approximation:

$$\frac{\sum_k \left(\hat{s}_k - s_k\right)^2}{\sum_n \left(\hat{f}_n - f_n\right)^2} \sim \frac{N_k}{N_n}, \tag{2.10}$$

where $N_k \equiv \sum_k$ and $N_n \equiv \sum_n$. This permits the following approximation of the ratio w:

$$w = \frac{\sum_k \Delta^{-2} \left(\hat{s}_k - s_k\right)^2}{\sum_n \sigma_n^{-2} \left(\hat{f}_n - f_n\right)^2} \sim \frac{N_k/\Delta^2}{N_n/\bar{\sigma}^2}, \tag{2.11}$$

where $\bar{\sigma} = \sum_n \sigma_n / N_n$. Then, for a desired ratio w, the requisite Δ can be approximated:

$$\Delta \simeq \bar{\sigma} \sqrt{\frac{N_k}{w N_n}}. \tag{2.12}$$

In practice, we have found that $w = \mathcal{O}(1)$ works well. The results of the training are relatively robust to the selection of w, in that changing w by, for example, a factor of 2 yields similar results.

2.4 Data

We collect a set of galaxy spectroscopic redshifts, paired with broadband photometry, from various surveys to test our training algorithm. Our set consists of 102,476 galaxies with redshifts $z < 4.54$ and i-band magnitudes[1] in the range $13.8 < i < 25.7$. For all surveys, we use galaxies with highly reliable spec-z's, photometry in one of the i-bands, and photometry in at least three bands with signal-to-noise ratio SNR > 20. The entire data set is summarized in Table 2.1, the filters used to measure the photometry are listed in Table 2.2, and the redshift distributions are shown in Figure 2.2.

2.4.1 zCOSMOS-bright

zCOSMOS [176] is a redshift survey of 1.7 deg^2 of the COSMOS field, conducted with the VIMOS spectrograph mounted on the European Southern Observatory's (ESO) Very Large Telescope (VLT). The survey is divided into two parts, *bright* and *deep*. We make use of the former, consisting of approximately 20,000 galaxies with redshifts $z < 1.2$. We use galaxies recommended in the ESO data release description[2], determined to have 99% spectroscopic verification (i.e. `zflag` = 3.x, 4.x, 2.5, 2.4, 1.5, 9.5, 9.3, 18.5, 18.3).

The zCOSMOS redshifts are matched to photometry from O. Ilbert et al. [140]. The photometry is measured from the ultraviolent to the near-infrared in 11 broadband filters: *NUV* on GALEX [197], u and i on CFHT-Megacam, B and V on CFHT-CFH12k, g^+, r^+, i^+, and z^+ on Subaru, and J on UKIRT. The final set consists of 14,298 galaxies with redshifts $z < 2.52$ and i-band magnitudes in the range $16.9 < i < 24.2$.

---

[1]The i-band magnitudes quoted in this section denote the magnitude in one of i, i_2, I, or i^+ as listed in Table 2.2. For galaxies with photometry in multiple i-bands, the magnitude used is the first to appear in that list.

[2]https://www.eso.org/sci/observing/phase3/data_releases/zcosmos_dr3_b2.pdf

Data Set	$N_{\rm gal}$	$f_{\rm gal}$	$z_{\rm mean}$	$z_{\rm max}$	i-band range	$i_{\rm mean}$	$\bar{\sigma}_i$
zCOSMOS	14298	0.14	0.57	2.52	$16.87 \leq i \leq 24.18$	21.19	0.022
VVDS	6915	0.07	0.67	4.54	$13.84 \leq i \leq 24.97$	20.86	0.014
VIPERS	69415	0.68	0.70	2.15	$17.66 \leq i \leq 23.08$	21.38	0.017
DEEP2/3	10695	0.10	0.71	1.91	$15.30 \leq i \leq 25.36$	21.42	0.020
3D-HST	1153	0.01	1.46	3.32	$19.10 \leq i \leq 25.74$	23.56	0.027
Training	81980	0.80	0.69	4.54	$13.84 \leq i \leq 25.74$	21.32	0.018
Test	20496	0.20	0.69	3.61	$16.46 \leq i \leq 25.69$	21.34	0.018
Total	102476	1.00	0.69	4.54	$13.84 \leq i \leq 25.74$	21.33	0.018

Table 2.1: Summary of the Spectrophotometric Data Sets. $N_{\rm gal}$ is the total number of galaxies in the set, $f_{\rm gal}$ is the fraction of all galaxies in the set, and $\bar{\sigma}_i$ is the mean fractional flux error in the i-band. Links to these catalogs are as follows: zCOSMOS: http://cesam.lam.fr/hstcosmos/; VVDS: https://cesam.lam.fr/vvds/index.php; VIPERS: http://vipers.inaf.it:8080/; DEEP2/3: http://d-scholarship.pitt.edu/36064/; 3D-HST: http://d-scholarship.pitt.edu/36064/.

2.4.2 VVDS

The VIMOS VLT Deep Survey (VVDS, O. Le Fèvre et al. 171) is a redshift survey consisting of three component surveys: *Wide*, *Deep*, and *Ultra-Deep*. The Wide survey covers 8.7 deg^2, with approximately 25,000 galaxies in the range $17.5 < i < 22.5$; the Deep survey covers 0.74 deg^2, with approximately 11,000 galaxies in the range $17.5 < i < 24$; the Ultra-Deep survey covers 512 arcmin^2, with approximately 900 galaxies in the range $23 < i < 24.75$. We use redshifts with quality flags 3 and 4, indicating a 98% spec-z confidence. The photometry was measured in nine filters: u, g, r, i, z on CFHT-Megacam [139] and B, V, R, I on CFHT-CFH12k [170]. The final set contains 6,915 galaxies out to redshifts $z < 4.5$, with magnitudes $13.8 < i < 25.0$.

Filter	Telescope	Instrument	λ_0	$W_{\rm eff}$
NUV	GALEX		2343.1	767.3
u	CFHT	Megacam	3817.7	525.4
B	CFHT	CFH12k	4342.5	873.6
B_J	Subaru	Suprime	4478.4	763.9
g^+	Subaru	Suprime	4808.5	1043.1
g	CHFT	Megacam	4899.9	1293.8
V	CFHT	CFH12k	5393.7	882.7
V_J	Subaru	Suprime	5493.0	862.4
r	CHFT	Megacam	6278.2	1120.2
r^+	Subaru	Suprime	6314.8	1211.4
R	CFHT	CFH12k	6603.5	1138.5
i_2	CHFT	Megacam	7584.5	1409.4
i	CHFT	Megacam	7676.6	1307.6
i^+	Subaru	Suprime	7709.1	1361.7
I	CFHT	CFH12k	8277.3	1816.7
z	CHFT	Megacam	8857.6	1040.1
z^+	Subaru	Suprime	9054.5	1012.3
Y	Subaru	Suprime	10216.0	996.2
J	UKIRT	WFCAM	12508.5	1476.8

Table 2.2: List of the Broadband Filters. Mean wavelength, $\lambda_0 = \int \lambda R(\lambda) d\lambda$, and effective width, $W_{\rm eff} = \mathrm{Max}[R(\lambda)]^{-1}$, are given in angstroms. Filters are listed in order of increasing λ_0. The response functions for each filter were obtained from the Spanish Virtual Observatory (SVO) Filter Profile Service.

2.4.3 VIPERS

The VIMOS Public Extragalactic Redshift Survey (VIPERS, M. Scodeggio et al. 272) is a dense, large-volume redshift survey focusing on redshifts $0.5 < z < 1.2$. We use VIPERS galaxies with spec-z's reliable at the 95% confidence level (`zflag` = 2.X, 3.X, 4.X), and with `photoMask` and `spectroMask` = 1. The redshifts are matched to photometry measured in *NUV* on GALEX [197], and u, g, r, i_2, i, z on CFHT-Megacam[3] [139]. The final set contains 71,951 galaxies with redshifts $z < 2.15$ and magnitudes $17.7 < i < 23.3$.

2.4.4 DEEP2 and DEEP3

DEEP2 and DEEP3 are redshift surveys conducted with the DEIMOS spectrograph on the Keck 2 telescope. DEEP2 [219] consists of four fields; we use galaxies from the first field in the Extended Groth Strip (EGS), which had no redshift pre-selection. DEEP3 [65] expanded on the DEEP2 survey of the EGS. Redshifts from these surveys are matched with aperture-corrected photometry provided by R. Zhou et al. [323]. We use galaxies with CFHTLS flag 0, SExtractor flags less than 4 in every band, and redshift quality flag ≥ 3. Photometry was measured in u, g, r, i_2, i, z on CFHT-Megacam[3] and Y on Subaru [207]. The final set contains 10,695 galaxies with redshifts $z < 1.91$ and magnitudes $15.3 < i < 25.74$.

2.4.5 3D-HST

In addition to the spectroscopic surveys above, we include grism redshifts from the 3D-HST survey [208, 219]. Redshifts for this survey were analyzed and matched with aperature-corrected photometry by R. Zhou et al. [323]. We select the galaxies with CFHTLS flag 0, SExtractor flags less than 4 in every band, and the flag `use_zgrism1` = 1. For galaxies in both the DEEP2/3 and 3D-HST sets, we use DEEP2/3 redshifts instead. Photometry

[3]The i_2 band is the replacement to the Megacam i-band installed in 2007. This filter is named y in the CFHTLS catalogues [139], but we follow R. Zhou et al. [323] in naming it i_2 to avoid confusion with the longer y bands used in Subaru and LSST.

was measured in u, g, r, i_2, i, z on CFHT-Megacam and Y on Subaru. After these cuts, the 3D-HST set contains 1,153 galaxies with redshifts $z < 3.32$ and magnitudes $23.6 < i < 25.7$.

2.5 Application to Data

Using the training algorithm described in Section 2.3, we will learn galaxy SED templates directly from the broadband photometry described in Section 2.4. We divide the data set into a training and test set, consisting of random 80% and 20% samples respectively of the entire data set. The training set will be used to train the SED templates, while the test set will be used to evaluate the learned templates via photo-z estimation (see Section 2.6). The training set consists of 81,980 galaxies, with mean redshift $z_{\rm mean} = 0.69$, max redshift $z_{\rm max} = 4.54$, and magnitudes $13.8 < i < 25.7$. A full summary of the set can be seen in Table 2.1, and the redshift distribution can be seen in Figure 2.2.

Eight naive templates were chosen to represent the underlying SED shapes of the photometry set according to the principles described at the end of Section 2.3.1. We chose the number eight to allow a direct comparison to the standard template set described below. They are "naive" because they are simply chosen by eye to roughly divide the photometry into groups by spectral shape, but otherwise are not based on any theoretical models or observed SED's. Each of the naive templates is a log-normal function,

$$S(\lambda) \propto \frac{1}{\lambda} \exp\left[-\frac{1}{2\eta^2}\left(\ln\frac{\lambda}{\text{mode}(\lambda)} - \eta^2\right)^2\right], \tag{2.13}$$

normalized at $\lambda = 5000$ Å, with mode(λ) in the range 1000 to 5500 Å and η in the range 0.35 to 0.9. The templates extend to 15000 Å with 100 Å resolution. These eight templates (hereafter N8) can be seen together with with their original training sets in Figure 2.3.

The training algorithm with $w = 0.5$ is applied to the N8 templates. The convergence of the templates is evaluated via the weighted mean square error,

$$\text{wMSE} = \sum_n \frac{1}{\sigma_n^2}(\hat{f}_n(\{\hat{s}_k\}) - f_n)^2. \tag{2.14}$$

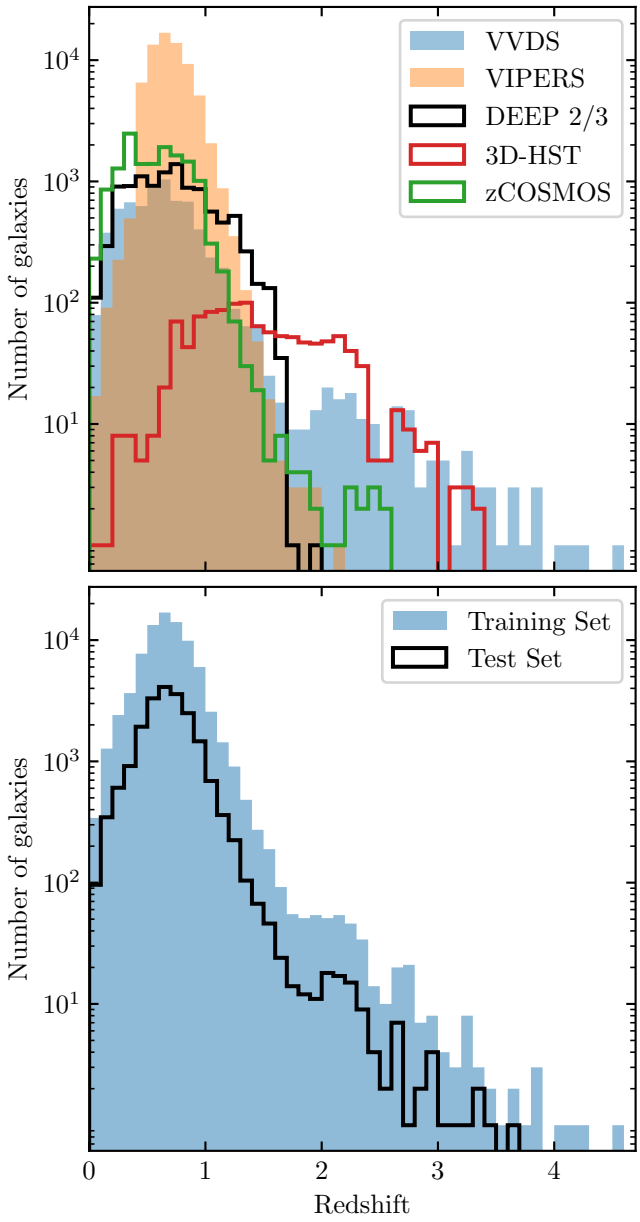

Figure 2.2: Redshift distribution of the galaxy surveys. The top panel shows the distributions of each of the constituent surveys. The bottom panel shows the redshift distributions of the training and test sets used for template training and photo-z estimation respectively.

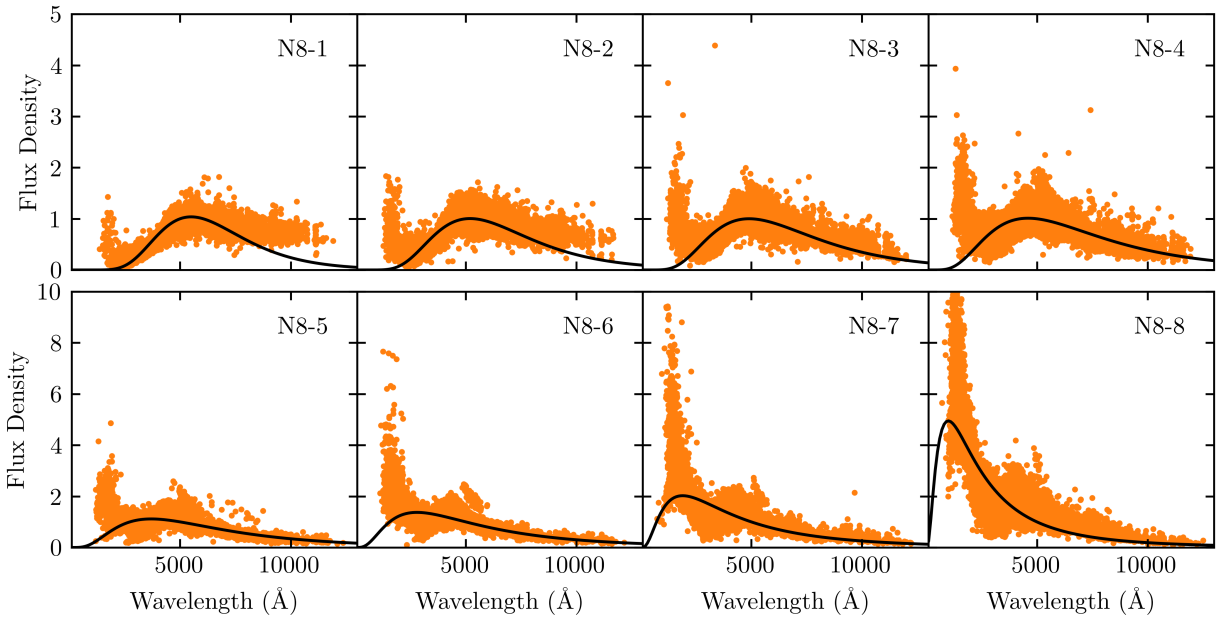

Figure 2.3: The untrained N8 templates (black lines) with their corresponding photometry sets (orange points), generated with the algorithm described in Section 2.3.1. N8-1 is the reddest template, with each successive template getting bluer.

Each template is perturbed until the change in wMSE is less than 3%, which was chosen empirically to balance sufficient template reconstruction and the algorithm's runtime. When every template has converged to its current photometry set, new photometry sets are generated. Only those templates whose new photometry sets result in a greater than 3% change in wMSE resume perturbation with their new sets. This process is iterated until no template has a new photometry set that results in a greater than 3% change in wMSE. This indicates that the photometry is sorted into distinct sets, and that further perturbation is unlikely to improve the photometry-matching results.

The progress of the training algorithm is shown in Figure 2.4 for the template N8-1. The left panel shows the progress of the perturbation algorithm as it deforms the originally smooth N8-1 template to better match the colors of the matched photometry sets. In particular, N8-1 becomes redder and acquires higher resolution structure, which will be discussed below. The middle panel shows the wMSE and the right panel shows the fractional change in the wMSE throughout the training. Orange points indicate values after a photometry-matching stage, and blue points indicate values after a perturbation. You can see that the wMSE drops as the template is perturbed, and perturbation continues until the magnitude of the fractional change in wMSE drops below 0.03, indicated by the dotted black lines in the right panel. Once this occurs, new photometry is matched, resulting in an increase in wMSE. This process is iterated, with fewer and fewer perturbations needed per iteration. Eventually, all of the points are orange, indicating that after each new photometry matching, N8-1 is not perturbed, as it already sufficiently matches its photometry set.

The training continues for 12 rounds, and takes approximately 15 minutes. The final results for the N8 templates can be seen in Figure 2.5. The templates are now a much better match to the photometry and more closely resemble physical galaxy spectra. Most of the templates have a Balmer Break at 4000 Å, although this was essentially already present in the initial templates. In addition, there are now emission and absorption lines visible in the spectra at a much higher resolution than the broadband filters used for the photometry (some of which are labeled with gray lines in Figure 2.5). Template N8-1 displays Mg and Na

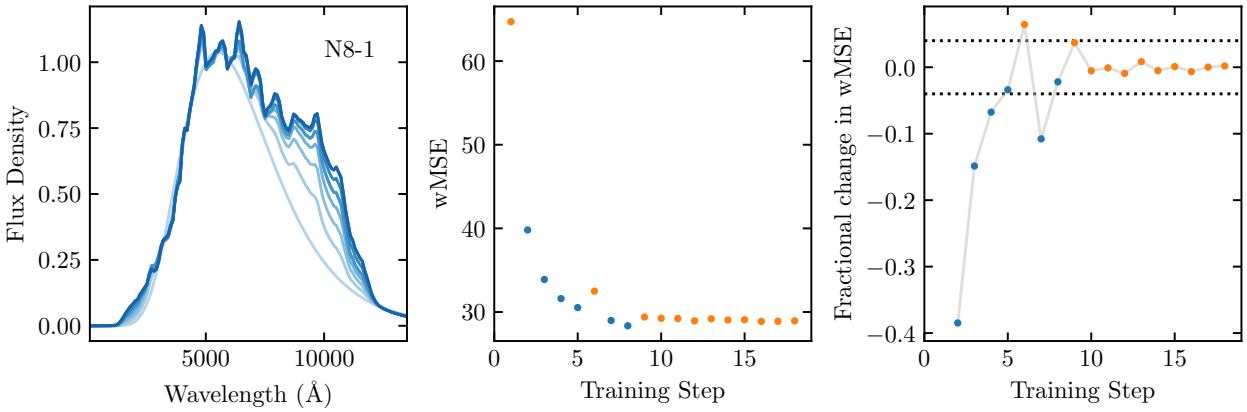

Figure 2.4: Training of N8-1. Left: the initial (light blue) N8-1 template is iteratively perturbed to better represent the colors of its photometry set. The final (dark blue) template is redder and has more structure. Middle: wMSE of the N8-1 template throughout the training process. Orange points represent the wMSE after a photometry matching stage, while blue points represent the wMSE after a perturbation. Right: fractional change in the wMSE. Orange points represent the fractional change due to a new photometry matching stage, while blue points represent a fractional change due to a perturbation. The dotted black lines show the ± 0.03 cutoff. When a perturbation results in a fractional change of magnitude less than 0.03, perturbation is halted and new photometry is matched. After the sixth photometry match, the template is not perturbed because it already sufficiently matches the photometry.

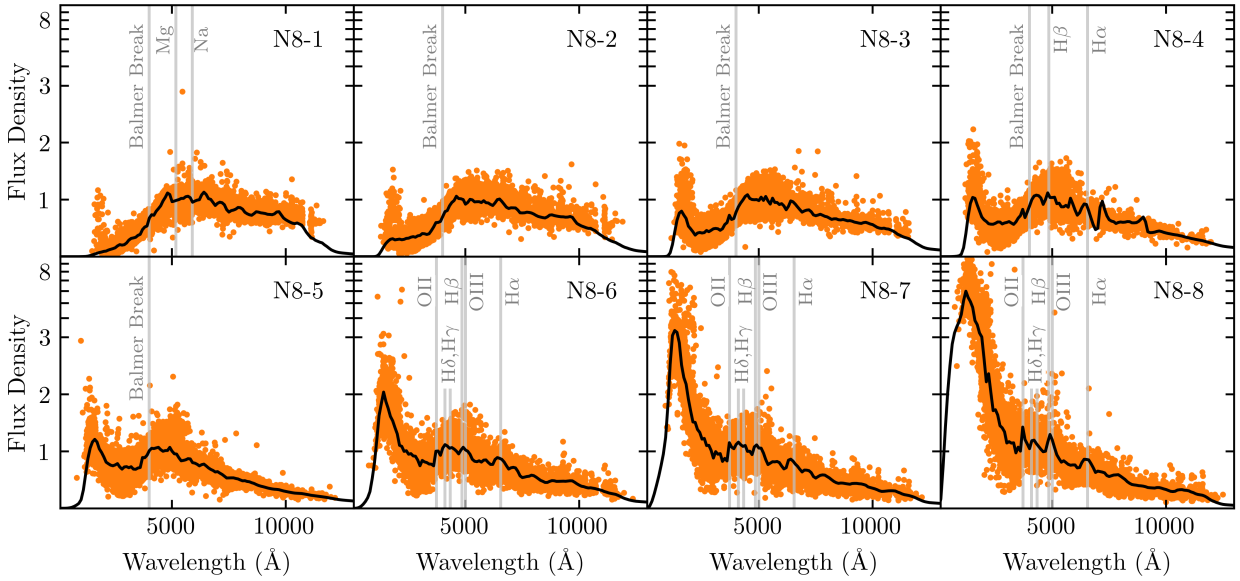

Figure 2.5: The trained N8 templates (black lines) with their final photometry sets (orange points). N8-1 is the reddest template, with each successive template getting bluer. The templates now more closely resemble physical galaxy spectra, and have acquired structure at a higher resolution than the broadband templates. The Balmer break, Mg and Na absorption lines, and Hα, Hβ, Hγ, Hδ, OII, and OIII emission lines are labeled in gray.

absorption lines and template N8-4 contains the beginnings of $H\alpha$ and $H\beta$ emission lines. Templates N8-6, N8-7, and N8-8 contain what appear to be Hα, Hβ, Hγ, Hδ, OII, and OIII emission lines (see Section 2.5.1 for more analysis). The emergence of these high resolution features from a large ensemble of low resolution data is the one of the defining features of this method.

In addition to these eight templates, we double the template number and train a set of 16 templates, in order to demonstrate the algorithm's ability to reconstruct templates with a more gradual transition of the colors from red to blue. This set (hereafter N16) was drawn from the same range of parameters for the log-normal function, and trained for 50 minutes over 26 rounds. The results of the training can be seen in Figure 2.6. These results closely

resemble the N8 results, with the same spectral features emerging. However, the N16 set shows a more gradual transition in color.

In addition to starting from naive templates, one can start with templates derived from spectral synthesis models or observations of local galaxy spectra [39, 74]. Here we apply the training algorithm to a standard set of SED templates commonly used for photo-z estimation (e.g. `BPZ`, see Section 2.6.1). This set (hereafter CWW+SB4) consists of four templates from G. D. Coleman et al. [60] and two starburst templates from A. L. Kinney et al. [162], the latter of which were added to account for faint blue galaxies in the HDF-N. These six templates were recalibrated by N. Benitez et al. [22] to correct for systematic differences between the observed and predicted galaxy colors in the HDF-N and other spectroscopic catalogs. In addition to these six, CWW+SB4 contains two synthetic starburst templates from G. Bruzual & S. Charlot [37], added by D. Coe et al. [59] to account for even bluer galaxies in the UDF.

The CWW+SB4 templates were trained with $w = 2$ for 46 minutes over 32 iterations. The results of the training can be seen in Figure 2.7. The original templates are plotted in blue, with the trained templates plotted in black, along with the final photometry sets in orange. You can see that the El and Sbc templates have barely been altered. The remaining templates have all systematically become redder. The high resolution structure that was originally present in the Im, SB3, and SB2 templates have been decreased in magnitude, while additional structure has been added to the simulated 25Myr and 5Myr templates what were originally smooth. These new features have been labeled in gray.

2.5.1 Reconstructing Spectral Lines

The template training algorithm allows the reconstruction of high resolution spectral features from low resolution photometry due to the oversampling of the underlying SED templates. This includes the emergence of spectral lines in many of the templates (c.f. Figures 2.5, 2.6, and 2.7). Knowledge of these lines allows us to perform post-processing of the learned templates to deconvolve the lines from the broadband filters. Here we perform a simple post-

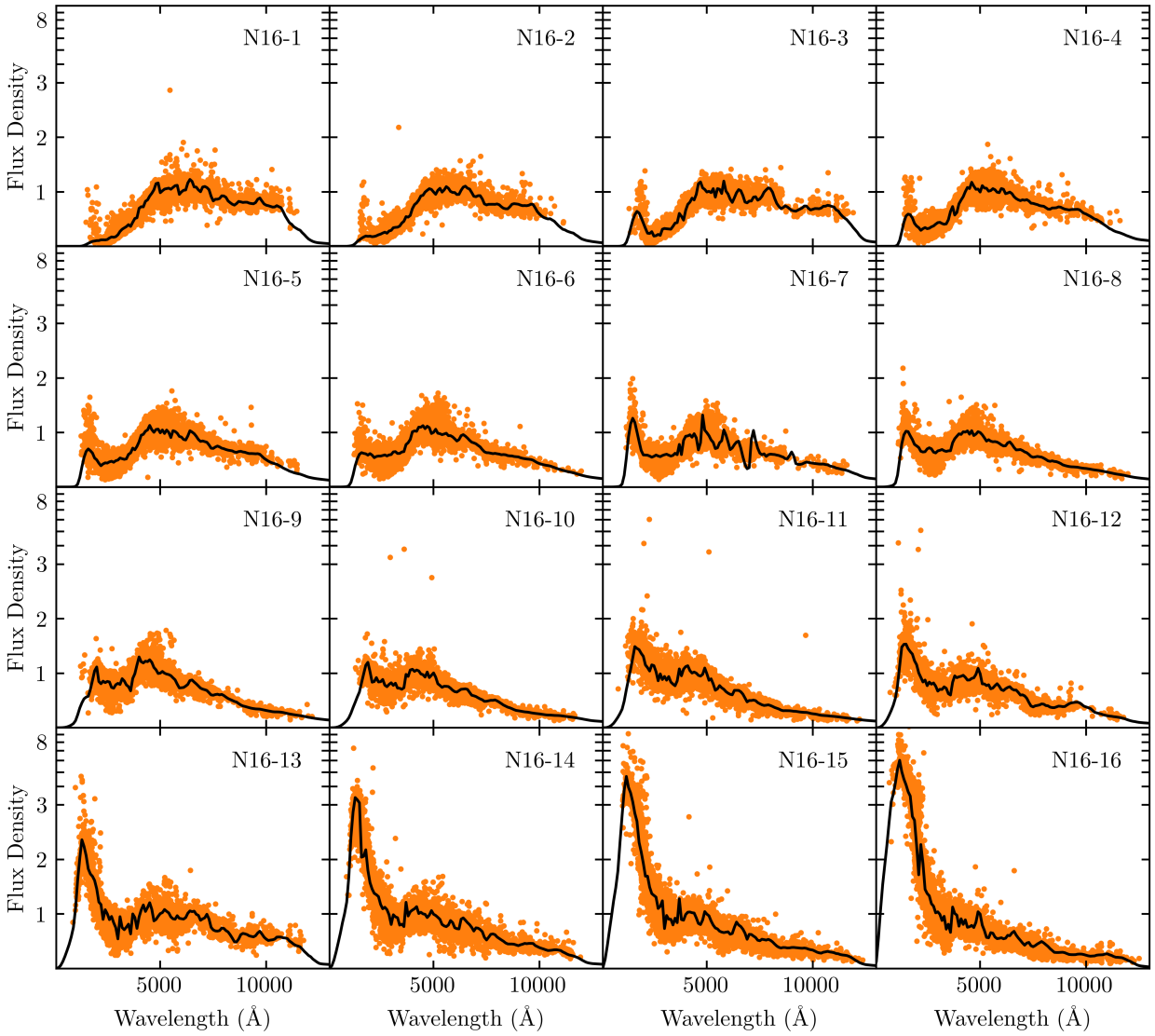

Figure 2.6: The trained N16 templates (black lines) with their final photometry sets (orange points). N16-1 is the reddest template, with each successive template getting bluer. These templates closely resemble the N8 templates and have show the same emerging spectral features (c.f. Figure 2.5), but consist of a more continuous transition from red to blue spectra.

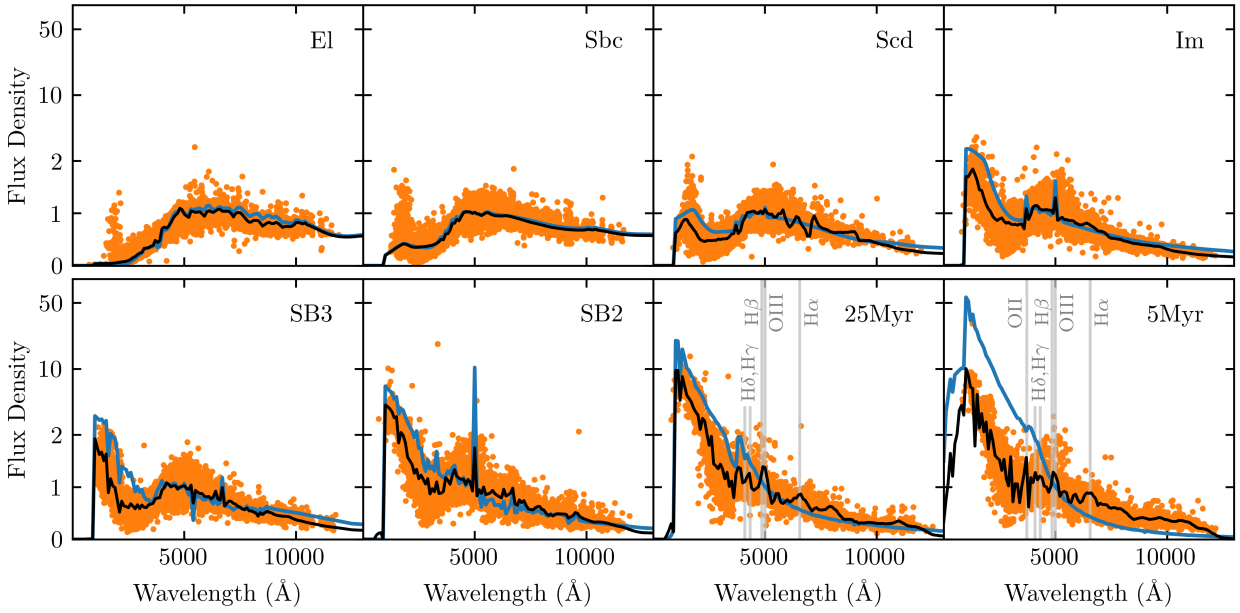

Figure 2.7: Result of training the CWW+SB4 templates. The original templates are in blue, the trained templates in black, and the final training sets are displayed as orange points. The 25Myr and 5Myr templates have acquired emission lines that were not present in the initial templates. These are labeled in gray.

processing of the N8-6, N8-7, and N8-8 templates to reconstruct the emission lines labeled in Figure 2.5. The templates are up-sampled to 10 Å and the continuum of each is linearly interpolated around the emission lines. The excess flux is attributed to the corresponding spectral lines. The flux of the Hβ line is impossible to distinguish from the OIII line in our templates because they are so close to one another. The same is true for the Hγ and Hδ lines. To overcome this difficulty, we use the Balmer decrements of 10^4K SDSS galaxies from B. Groves et al. [124]: Hα/Hβ = 2.86 and Hγ/Hδ = 1.81. We calculate the Hβ flux from Hα, and subtract this from the combined Hβ-OIII flux, and we calculate Hγ and Hδ from the combined Hγ-Hδ flux.

After calculating the flux of the emission lines, the final templates are built by adding Gaussians of equivalent amplitude and FWHM = 20 Å to the continuum. The templates

		N8-6		N8-7		N8-8	
Line	λ	r	W_λ	r	W_λ	r	W_λ
Hα	6563	2.86	132.7	2.86	103.3	2.86	115.2
Hβ	4861	1.00	32.9	1.00	26.4	1.00	30.3
Hγ	4340	1.18	36.5	1.31	31.6	1.28	37.1
Hδ	4102	0.65	19.6	0.72	16.7	0.71	20.7
OII	3727	2.04	58.1	1.27	32.0	0.74	24.4
OIII	5007	2.08	68.0	2.42	66.1	0.86	27.3

Table 2.3: Reconstructed Emission Lines. or each emission line, r is the amplitude relative to Hβ, and W_λ is the effective width in angstroms.

with the reconstructed spectral lines can be seen in Figure 2.8. For each line, we calculate the amplitude relative to Hβ, and the effective width, $W_\lambda = \int(1 - F_\lambda/F_0)d\lambda$, where F_λ is the total flux, and F_0 is the continuum flux. These values can be seen in Table 2.3. Note that the amplitudes of our reconstructed Hγ and Hδ lines relative to Hβ are approximately three times greater than those listed in B. Groves et al. [124].

2.6 Estimating Photo-z's

We evaluate the results of our template training algorithm by using our learned templates to estimate photo-z's for the test set of galaxies using the software package `BPZ` [21], and comparing the results to the spec-z's and the photo-z's estimated using the original CWW+SB4 templates. The test set consists of 20,496 galaxies (20% of the total set), with mean redshift $z_{\text{mean}} = 0.69$, max redshift $z_{\text{max}} = 3.61$, and magnitudes $13.8 < i < 25.7$. See Table 2.1 for a full summary and Figure 2.2 for the redshift distribution.

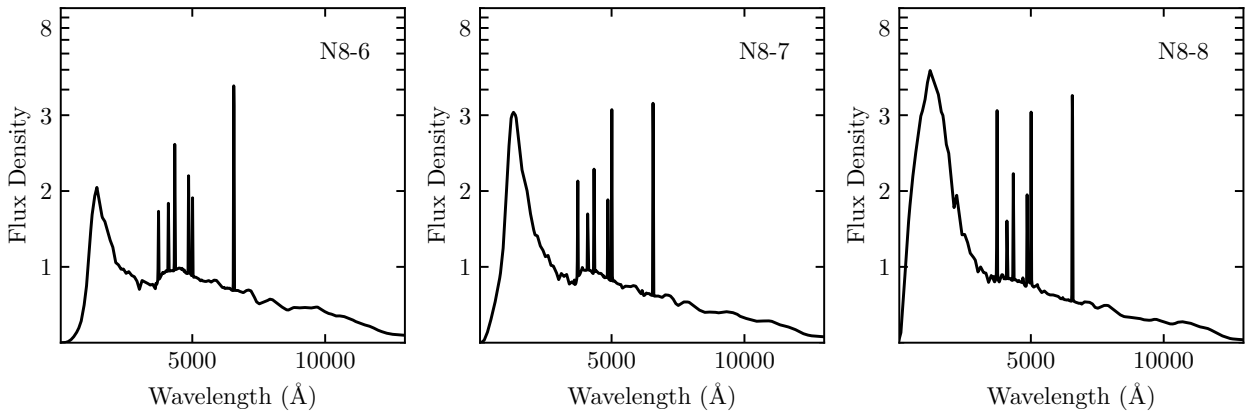

Figure 2.8: The N8-6, N8-7, and N8-8 templates with reconstructed emission lines (cf. Figure 2.5). The emission lines, left to right, are OII,Hδ, Hγ, Hβ, OIII, and Hα. The wavelengths, relative amplitudes, and effective widths of these lines are in Table 2.3.

2.6.1 Bayesian Photometric Redshifts

Bayesian Photometric Redshifts (`BPZ`; N. Benítez 21) is a template-based photo-z estimator. Template-based estimators take a set of SED templates, assumed to be spanning and exclusive, and calculate observed fluxes over a grid of redshift values. For each template, `BPZ` evaluates a χ^2 function at each redshift on the grid:

$$\chi^2(z, T, A) = \sum_n \frac{1}{\sigma_n^2}(A\,\hat{f}_n(z, T) - f_n)^2, \tag{2.15}$$

where T denotes the template, z denotes the redshift, A is a normalization, and $\hat{f}_n$, f_n, and σ_n denote the calculated flux, the observed flux, and the fractional error as in Equation 2.5. The sum over n is a sum over the filters for the set of observed fluxes. `BPZ` then evaluates the likelihood for producing the observed galaxy fluxes: $p(\{f_n\}|z, T) \propto \exp(-\chi^2/2)$. The redshift posterior is then calculated by marginalizing over the set of templates:

$$\begin{aligned} p(z|\{f_n\}, m_0) &= \sum_T p(z, T|\{f_n\}, m_0) \\ &\propto \sum_T p(z, T|m_0)\, p(\{f_n\}|z, T), \end{aligned} \tag{2.16}$$

where $p(z, T|m_0)$ is a prior over the apparent magnitude m_0. Work is underway to determine how best to use the full information encoded in the redshift posterior generated by `BPZ` and other photo-z codes (e.g. S. J. Schmidt et al. 268). In this work, however, only the mode of the posterior distribution is used to estimate the photo-z.

We use `BPZ`-v1.99.3[4] to estimate photo-z's. We turn off template interpolation by setting `INTERP=0`. For simplicity, we treat non-detections as non-observations. We use the various sets of SED templates described in Section 2.5, and use the prior described in the following section. All other settings were left as default.

2.6.2 Galaxy Magnitude Priors

Before estimating photo-z's with `BPZ`, we must first construct the magnitude priors, $p(z, T|m_0)$, calibrated to the galaxies in our training set. We separate the prior into two parts:

$$p(z, T|m_0) = p(T|m_0)\, p(z|T, m_0) \tag{2.17}$$

For the magnitude m_0, we use one of the i bands in the following order of priority: i, i_2, I, i^+. Instead of constructing a different prior for each template, we follow N. Benítez [21] in dividing our templates into three broad classifications: elliptical (El), spiral (Sp), or irregular/starburst (Im/SB). The CWW+SB4 templates are already classified under this scheme. We classify our new templates and each of the galaxies in the training set by assigning the classification of the CWW+SB4 template with the most similar colors, determined by minimizing the mean square error of the fluxes.

The N8 templates are determined to have one elliptical, four spiral, and three irregular/starburst galaxies; the N16 templates are determined to have two elliptical, eight spiral, and six irregular/starburst galaxies. The fraction of each classification as a function of magnitude for the training set galaxies is displayed in Figure 2.9.

---

[4] `http://www.stsci.edu/~dcoe/BPZ/`

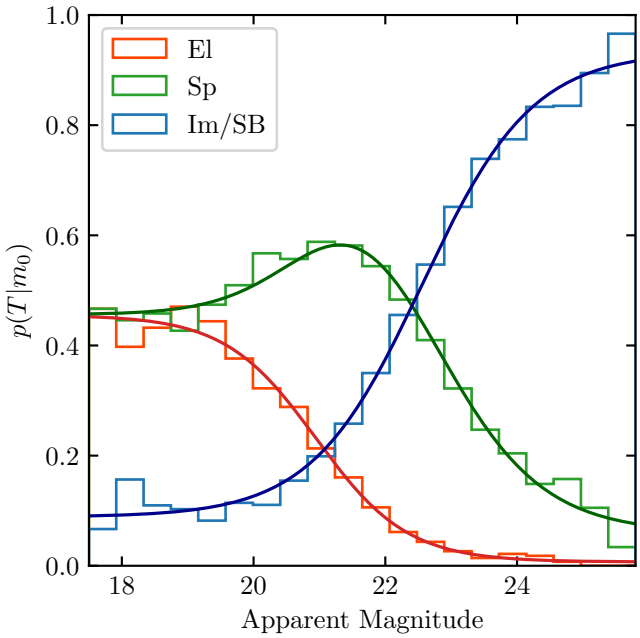

Figure 2.9: Fraction of each spectral class as a function of apparent magnitude. The histograms represent the fractions in the training set, and the curves are the spectral type priors fit to the data.

We assume that the El and Im/SB galaxies have spectral priors of the form

$$p(T|m_0) = \frac{L_T}{1 + e^{-\kappa_T(m_0 - m_T)}} + C_T, \tag{2.18}$$

while $p(\mathrm{Sp}|m_0) = 1 - p(\mathrm{El}|m_0) - p(\mathrm{Im/SB}|m_0)$. The values of $\{L_T, \kappa_T, m_T, C_T\}$ for the El and Im/SB galaxies are found by fitting to the distributions in Figure 2.9. All three priors are plotted in the same figure, and the parameter values are listed in Table 2.4.

For the redshift prior, we use Equations 23 and 24 from N. Benítez [21]:

$$p(z|T, m_0) = \frac{1}{N_T} \exp\left\{-\left(\frac{z}{Z_T}\right)^{\alpha_T}\right\}, \tag{2.19}$$

where the normalization is

$$N_T = \frac{{Z_T}^{\alpha_T+1}}{\alpha_T}\, \Gamma\left(\frac{\alpha_T + 1}{\alpha_T}\right), \tag{2.20}$$

Spectral Type	El	Sp	Im/SB
L_T	0.448 ± 0.017	...	0.845 ± 0.031
κ_T	-1.45 ± 0.16	...	1.20 ± 0.11
m_T	21.0 ± 0.1	...	22.6 ± 0.1
C_T	0.007 ± 0.009	...	0.089 ± 0.013
α_T	3.88 ± 0.04	3.40 ± 0.04	2.22 ± 0.03
z_{0T}	0.484 ± 0.003	0.493 ± 0.003	0.361 ± 0.009
k_T	0.119 ± 0.002	0.124 ± 0.002	0.130 ± 0.008

Table 2.4: Parameters for the BPZ spectral priors, $p(z, T|m_0)$.

and the "median" redshift Z_T is chosen to have the linear dependence

$$Z_T(m_0) = z_{0T} + k_T(m_0 - 20). \tag{2.21}$$

Equation 2.19 reproduces the exponential cutoff at high redshifts present in the training set, and can reasonably approximate any unimodal redshift distribution, from very narrow ($\alpha \gg 2$) to very broad ($\alpha \ll 1$). This flexibility reduces the bias introduced by the functional form of the prior [21]. The nine parameters $\{\alpha_T, z_{0T}, k_T\}$ are determined by maximizing the likelihood $L = \prod_i p(z_i|T_i, m_{0i})$, where the product is over the galaxies in the training set. The parameters and their bootstrapped uncertainties are listed in Table 2.4.

2.6.3 Photo-z Results

We estimate photo-z's for the test set galaxies using `BPZ` with the settings and priors described in the previous two sections. We used four template sets: the original CWW+SB4 templates, the trained CWW+SB4 templates, and the trained N8 and N16 templates.

`BPZ` provides two metrics for the photo-z estimates: `ODDS` and $\chi^2_{\rm mod}$. `ODDS` measures how narrowly peaked the posterior distribution $p(z|\{f_n\}, m_0)$ is around the estimated photo-z.

Galaxies with low ODDS have either broad redshift posteriors, or posteriors with multiple peaks. $\chi^2_{\rm mod}$ measures how well the best fit template at the predicted redshift matches the observed fluxes. For more about these metrics, see Section 4 of N. Benítez [21] and Section 4.3 of D. Coe et al. [59]. In this work, photo-z estimates with ODDS < 0.95 or $\chi^2_{\rm mod} > 1$ are excluded from the analysis, and the fraction excluded on this bases is reported as $f_{\rm cut}$.

To further evaluate the results of BPZ, we calculate the scatter, bias, and outlier fraction of the photo-z estimates. Photo-z estimates are known to be contaminated with a significant number of outliers. This is largely driven by a degeneracy wherein the 1000Å Lyman break in a high redshift galaxy spectrum has similar optical colors to the 4000Å Balmer break in a low redshift galaxy spectrum. BPZ attempts to break this degeneracy with the galaxy magnitude prior (i.e. galaxies with brighter apparent magnitudes are more likely to be at a lower redshift), yet there are still a large number of outliers.

To address this issue, we evaluate the statistics of the interquartile range (IQR) of the data, as these measures are robust to the presence of outliers. We follow M. L. Graham et al. [119] in introducing the quantity $\Delta z_{1+z} = (z_{spec} - z_{phot})/(1 + z_{phot})$. The numerator quantifies the photo-z error and the denominator compensates for the larger uncertainty at high redshifts. We define the scatter of the photo-z estimates, $\sigma_{\rm IQR}$, as the width of the IQR in Δz_{1+z}, divided by 1.349 to convert to the equivalent of a Gaussian standard deviation. We define the bias of the photo-z estimates as the mean value of Δz_{1+z} for galaxies within the IQR. The uncertainties of these two values are bootstrapped by calculating the values on 1000 random samples with replacement. Outliers are identified as photo-z's with $\Delta z_{1+z} > 3\sigma_{\rm IQR}$, and the fraction of outliers is reported as $f_{\rm out}$.

The photo-z results can be seen in Figure 2.10. The photo-z estimates that passed the cuts on ODDS and $\chi^2_{\rm mod}$ are displayed as points: the inliers in blue, the outliers in orange. The values of the photo-z statistics for each template set are printed in each panel. For all four template sets, the photo-z estimation is reasonably accurate for spec-z's $z < 1.5$. For higher redshifts, there appears to be a systematic bias towards higher photo-z's. Reduced photo-z accuracy is generally expected for spec-z's greater than 1.5, as the Balmer break leaves the

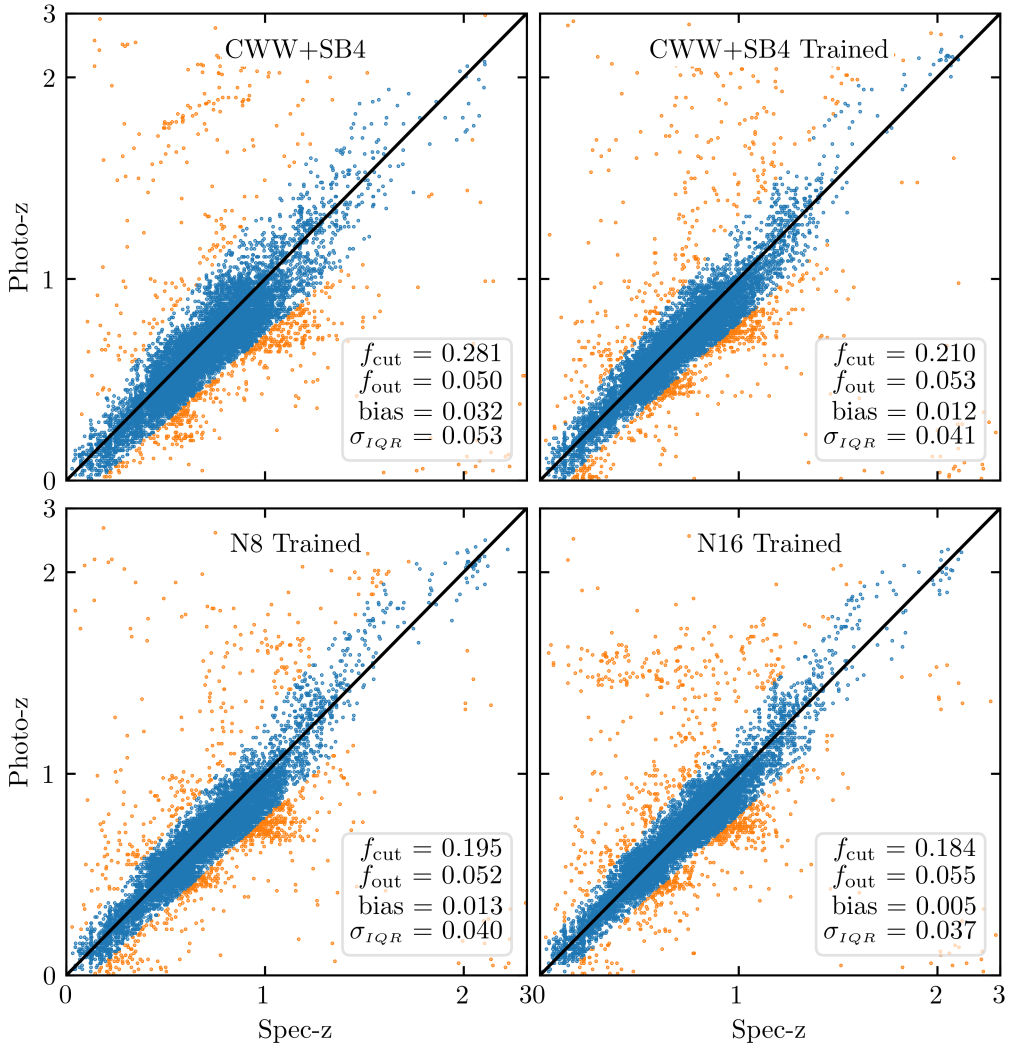

Figure 2.10: Results of photo-z estimation with `BPZ`, using the four different templates sets. Photo-z estimates are displayed as points: inliers are blue and outliers are orange. The black line represents perfect estimation (i.e. photo-z = spec-z). The statistics printed in each panel are for the entire data set.

optical bands at around $z = 1.4$ and the Lyman break does not enter the ultraviolet bands until $z = 2.5$.

For the CWW+SB4 templates, the training algorithm decreased the fraction of photo-z's cut by 25%, the bias by 63%, and the scatter by 23%, but did not improve the outlier fraction. We were able to achieve similar photo-z results using the trained N8 and N16 template sets, demonstrating that our training algorithm can be used to generate photo-z templates without any a priori information about galaxy spectra. Compared to the CWW+SB4 templates, N8 templates decreased $f_{\rm cut}$ by 31%, bias by 59%, and scatter by 25%. The N16 templates decreased $f_{\rm cut}$ by 35%, bias by 84%, and scatter by 30%. In all cases, the training algorithm decreases the fraction of bad photo-z's ($f_{\rm cut} + f_{\rm out}$), the bias, and the scatter.

Comparing the results for the N8 and N16 template sets indicate that increasing the number of templates can reduce the fraction cut, and the bias and scatter of the photo-z estimates. To further investigate this relationship, we calculate the photo-z statistics for a range of template numbers, the results of which are in Figure 2.11. We find that increasing the number of templates decreases the fraction cut and the bias, as well as slightly decreasing the scatter. The trend for outlier fraction is less clear.

The N20 set has $f_{\rm cut} = 0.188$ (a 33% decrease compared to CWW+SB4), $f_{\rm out} = 0.040$ (a 20% decrease), bias = 0.003 (a 91% decrease), and scatter = 0.039 (a 26% decrease).

The value of the metrics as a function of photo-z can be seen in Figure 2.12. In addition to the template sets plotted above, we add the N20 set. For comparison, plotted in gray are the LSST science requirements for the metrics as listed in the LSST Science Requirement Document (SRD; Ž. Ivezić & LSST Science Collaboration 145). The SRD lists the following minimum requirements to enable the envisioned LSST cosmological studies: root-mean-square error $< 0.02(1 + z_{\rm phot})$; $f_{\rm out} < 10\%$; average bias $< 0.003(1 + z_{\rm phot})$. The SRD lists these requirements for an $i < 25$, magnitude-limited sample of four billion galaxies from $0.3 < z < 3.0$. For comparison, our test set consists of 20,496 galaxies with $i < 25.7$, in the range $z < 3.6$, including 19,391 galaxies with $i < 25$, in the range $0.3 < z < 3.0$. In Figure 2.12, we show that for redshifts $0.3 < z < 1.2$ we are able to achieve an appropriate outlier

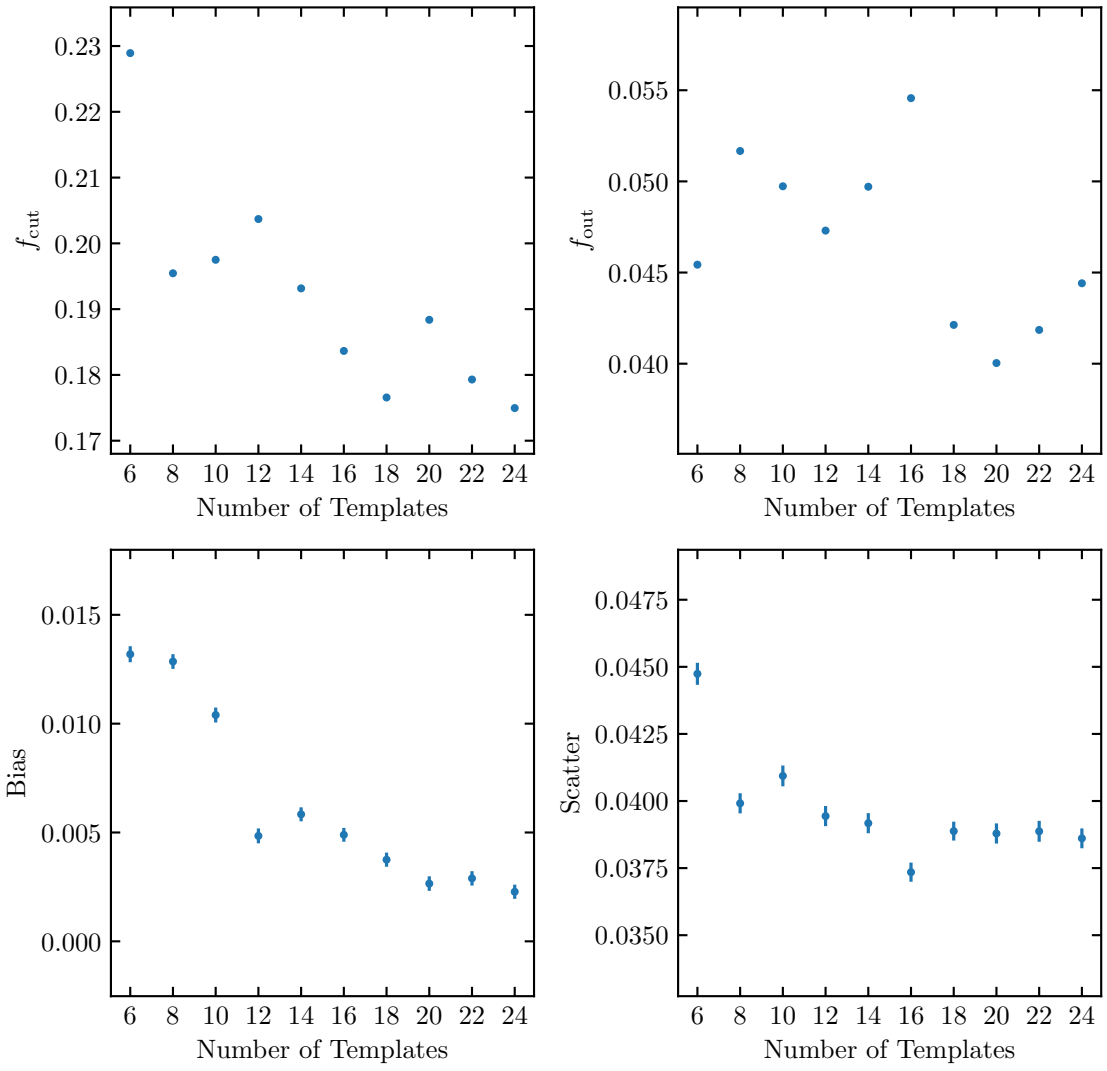

Figure 2.11: Photo-z statistics as a function of template number. Statistics are for the full redshift range.

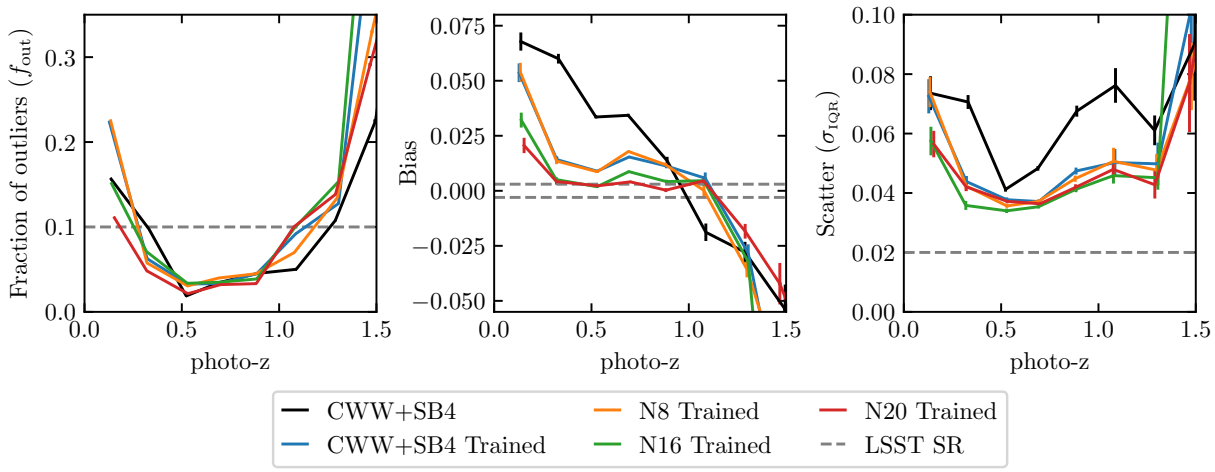

Figure 2.12: Photo-z metrics for the various template sets as a function of redshift bin. LSST science requirements are shown as dashed gray lines.

fraction, and that our training algorithm makes great progress on the bias, almost reaching the threshold required for LSST. We also make modest progress on the scatter, but reduction by another factor of two is still required. Beyond redshift $z = 1.2$, all of our metrics fail the LSST science requirements.

2.7 Discussion

In Section 2.3, we demonstrated that our training algorithm could learn galaxy SED templates from photometry at a high resolution relative to the filters used to make the observations. We are able to learn a set of templates over twice the size of the standard CWW+SB4, showing a smooth progression of galaxy colors from red to blue. The spectra contain relatively high resolution spectral features, and post-processing can further reconstruct emission and absorption lines. The bluer templates contain more structure as they represent star forming galaxies and thus have stronger emission lines. In addition, the bluer templates have a larger number of high-redshift galaxies compared to the red templates, which aids the reconstruction of high-resolution features. While the high-redshift galaxies number in

the hundreds instead of thousands, our results indicate that high-resolution features can be reliably reconstructed with only a few hundred galaxies.

Our method has a number of limitations. The success of our algorithm relies on the ability to generate a naive set of templates as a starting point that will reliably divide the photometry by the spectral type of the galaxy. This is relatively easy to accomplish for fewer than 20 templates, as was demonstrated by our simple photometry matching procedure and the log-normal templates we used. This is a strength of the algorithm as it is relatively robust to the starting templates. If, however, you wish to derive more than 20 templates from the photometry, care must be taken in the division of the photometry set to ensure there are sufficient galaxies in each subset to fully sample the entire wavelength range for the templates. In addition, the inherently discretized way in which we divide the photometry set stands in the way of generating a truly continuous set of SED templates. For a more continuous set of templates one might imagine taking two "adjacent" photometry sets, and assembling a photometry set "between" them by taking the bluer half on one set together with the redder half of the other. Equally we could construct a moving window that progressively subdivides a sample based on color (with galaxies allowed to be present in more than one subet).

Our data consists only of broadband photometry, however our algorithm would work equally well with narrow bands as well. Combining broadband and narrow band photometry would expand the data set and further constrain the templates. In particular, the addition of narrowband photometry should increase the resolution of spectral features recovered, and may allow one to resolve features such as the Hγ and Hδ emission lines that we had to treat as a single feature. One could also include bands from a wider range of wavelengths to increase the wavelength range over which the templates are constrained. We attempted to include fluxes from the K-bands included with the zCOSMOS and VIPERS catalogs to learn infrared wavelengths for the templates, but there appeared to be systematic calibration issues in the data that we could not resolve. There is evidence that the inclusion of near-infrared and near-ultraviolet photometry in photo-z estimation can reduce outliers and scatter by up to 50% each [120].

In addition, for the results presented here, we used only galaxy fluxes with SNR greater than 20. One can use galaxies with lower SNR if outlier fluxes are removed from the photometry sets before training (we had success using an Isolation Forest; F. T. Liu et al. 180, K. Ting et al. 297). However, lowering the SNR of the photometry generally reduces the resolution of the structure you can reconstruct.

The training algorithm itself could be made more sophisticated by restoring the wavelength dependence of the hyperparameter Δ_k. We also hope to move beyond an iterative regression approach into deep learning, perhaps using Generative Adversarial Networks (GANs; I. J. Goodfellow et al. 117).

When constructing the `BPZ` prior, we sorted our templates into broad spectral classes. In the N8 set, for example, we determined that one template was elliptical, four were spiral, and three were irregular/starburst. Each of our templates has approximately the same number of galaxies matched to it, and the photometry matched to the elliptical templates does not display more variance than the photometry matched to other templates. These observations indicate that our data set contains a larger number of spiral and irregular/starburst galaxies than elliptical galaxies, rather than suggesting that the space of elliptical galaxy spectra is less finely sampled. For this reason, we do not expect the imbalance of the template number in each class to have a large impact on the photo-z quality, but nevertheless note that a more sophisticated prior could be constructed without relying on this broad classification scheme which may provide better redshift estimates.

We found in Section 2.6.3 that our training algorithm can improve the bias and scatter of photo-z estimates. We found that increasing the number of templates enhances these improvements, with the best results for 20 templates. As mentioned above, with our current method for generating photometry sets, we struggle to reliably reconstruct more than 20 templates, so whether these benefits continue to decrease with template number is unknown.

We can compare our method for generating more SED templates with `BPZ`'s method of linearly interpolating between templates. N8 with `INTERP=2` generates 22 total templates. Table 2.5 compares the photo-z results using these templates with the results using 22 tem-

	INTERP	Total N	$f_{\rm cut}$	$f_{\rm out}$	Bias	Scatter
N8	0	8	0.228	0.058	0.014	0.040
N8	2	22	0.209	0.060	0.012	0.037
N22	0	22	0.214	0.045	0.004	0.039

Table 2.5: Comparison to BPZ Interpolation. Total N is the total number of SED templates in the set, including those interpolated by `BPZ`. Statistics quoted are for the full redshift range.

plates learned from the photometry with `INTERP=0`. It is clear that, as far as $f_{\rm out}$ and bias, our method for generating extra templates is superior to the linear interpolation used by `BPZ`.

The photo-z estimation with our learned template sets outperforms the results of the standard CWW+SB4 templates, however, more work needs to be done to reach the requirements set for LSST, especially for redshifts $z > 1$. Templates can be trained for LSST science using the substantial overlap of LSST photometry with the eBoss [77] and Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 82) surveys which will provide hundreds of thousands of spec-z's for LSST photo-z training and calibration [220, 267].

Our training method can be extended to other domains (e.g. stellar spectral reconstruction) where you can take a large set of incomplete data, segment that data into classes, and treat the set of unique observations in each class as an ensemble of observations of some class archetype, and thereby reconstruct more complete information. We plan to adapt the method to reconstruct supernova lightcurves from supernova photometry.

2.8 Conclusions

We have shown that galaxy SED templates can be learned directly from a data set of broadband photometry. Large sets of photometry at various redshifts can be leveraged to recon-

struct high resolution features, such as the Hα, Hβ, Hγ, Hδ, OII, and OIII emission lines, as well as Na and Mg absorption lines. Simple post processing can further improve the resolution of these reconstructed lines. The number of templates learned is variable and can be increased to more continuously sample the space of galaxy spectra and to improve photo-z results.

We used our templates to estimate photo-z's for a test set of galaxies using `BPZ`. We found that training the standard set of templates that comes with `BPZ` decreases the fraction of bad photo-z's by 21%, the bias by 63% and the scatter by 23%. Our own trained naive templates yielded better results. We learned a set of 20 templates from the data that reduced the fraction of bad photo-z's by 31%, the bias by 91%, and the scatter by 26%. These derived templates outperform the interpolated spectra used by BPZ. The improvements in bias are almost sufficient to meet the requirements set for LSST, but another reduction by a factor of two is needed for the scatter.

The templates derived with our training algorithm demonstrate that accurate galaxy spectra can be learned from broadband photometry. Our SED's could potentially be used for applications other than photo-z's, and our learning algorithm can be extended to other applications, such as learning supernova lightcurves from photometry.

Our derived templates and the code used to produce these results are publicly available in a dedicated Github repository: https://github.com/dirac-institute/photoz_template_learning.

Chapter 3

PROBABILISTIC FORWARD MODELING OF GALAXY CATALOGS WITH NORMALIZING FLOWS

Photo-z's are a key ingredient to cosmology with photometric surveys. Photo-z estimation is difficult, however, and biases from photo-z estimation can greatly bias the cosmological inferences that rely on them [218]. This motivates the LSST Dark Energy Science Collaboration (DESC) to perform extensive simulation studies to understand what kinds of biases our photo-z estimation pipelines are susceptible to, and how these biases will propagate to cosmology inferences. It has been difficult, however, to validate photo-z *posteriors* because spectroscopic data sets and most simulations do not have a natural notion of a *true* photo-z posterior to which the estimated posterior can be compared. Futhermore, attempts to circumnavigate this deficiency were found by S. J. Schmidt et al. [268] to be inadequate and misleading. The following chapter, originally published in *The Astronomical Journal 168, 80 (2024)*, addresses this problem by introducing a new simulation framework using the *normalizing flow* deep learning architecture. This work also resulted in the publication of the Python packages `pzflow`[1] and `PhotErr`[2], which have already seen wide use in the DESC for a diverse set of studies including optimizing survey strategy [195], quantifying survey non-uniformity [127], simulating correlated host galaxies for type Ia supernovae [184], and augmenting spectroscopic training sets with simulated data [210].

---

[1] https://github.com/jfcrenshaw/pzflow

[2] https://github.com/jfcrenshaw/photerr

3.1 Abstract

Evaluating the accuracy and calibration of the redshift posteriors produced by photometric redshift (photo-z) estimators is vital for enabling precision cosmology and extragalactic astrophysics with modern wide-field photometric surveys. Evaluating photo-z posteriors on a per-galaxy basis is difficult, however, as real galaxies have a true redshift but not a true redshift posterior. We introduce PZFlow, a Python package for the probabilistic forward modeling of galaxy catalogs with normalizing flows. For catalogs simulated with PZFlow, there is a natural notion of "true" redshift posteriors that can be used for photo-z validation. We use PZFlow to simulate a photometric galaxy catalog where each galaxy has a redshift, noisy photometry, shape information, and a true redshift posterior. We also demonstrate the use of an ensemble of normalizing flows for photo-z estimation. We discuss how PZFlow will be used to validate the photo-z estimation pipeline of the Dark Energy Science Collaboration (DESC), and the wider applicability of PZFlow for statistical modeling of any tabular data.

3.2 Introduction

Photometric redshift (photo-z) estimation is necessary for the study of cosmology and galaxy evolution with the huge galaxy catalogs generated by modern wide-field photometric surveys like the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST; Ž. Ivezić et al. 146). The colors of galaxies in these surveys are used to estimate posterior distributions over the possible redshift of each galaxy. These posteriors are often multimodal, reflecting degeneracies in redshift and galaxy spectral type. Guaranteeing the accuracy and calibration of these posteriors, and understanding their biases, is vital for enabling precision cosmology and studies of galaxy evolution [218, 291].

There have been many studies evaluating the performance of photo-z estimation by comparing estimates to known true redshifts (e.g., H. Hildebrandt et al. 132, C. Sánchez et al. 261, M. L. Graham et al. 119), however to meet the needs of modern surveys, we must evaluate the accuracy of the full redshift posteriors generated for these galaxies. In other words,

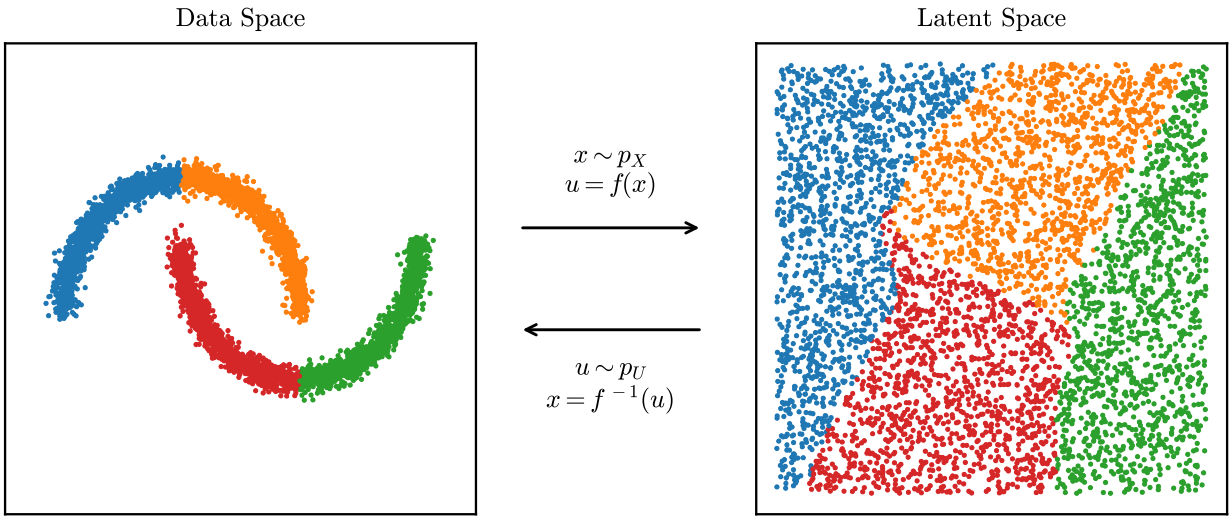

Figure 3.1: A normalizing flow demonstrated on the two moons data set from scikit-learn. The two moons data on the left is mapped onto a two dimensional uniform distribution by the bijection f. The data are colored by quadrant to visualize their image in the latent space. You can sample the data distribution by sampling from the uniform distribution, and using f^{-1} to map the samples back to the data space.

not only must we verify that photo-z estimators assign a high posterior probability to the true redshift, but also that the posteriors account for all the degeneracies inherent in photo-z estimation, and that the distributions are neither too narrow nor too broad. This kind of validation is difficult, however, because individual galaxies do not have a "true" redshift posterior to which you can compare the estimated posterior. S. J. Schmidt et al. [268] studied whether these questions could be answered using ensemble-level comparisons of photo-z posteriors and true redshifts, but found that these metrics could be fooled by a pathological photo-z estimator that ignored galaxy colors altogether. This raised the need for new methodologies that enable the evaluation of photo-z posterior estimates at the per-galaxy level.

One way to enable posterior validation at the per-galaxy level is to simulate catalogs

by sampling galaxies from a probabilistic model for which you have direct access to the underlying probability distribution. Then, for each galaxy, you can calculate the posterior over redshift directly from the same model that generated the galaxy. This provides a "true" redshift posterior for each simulated galaxy to which you can compare the posteriors from photo-z estimators, enabling the validation of individual photo-z posterior estimates.

Normalizing flows are a generative model of this variety. Normalizing flows model complex, high-dimensional probability distributions using deep neural networks that learn an invertible mapping between the complicated data distribution and a more simple distribution, known as the latent distribution. This allows you to sample from and calculate probabilities with respect to the latent distribution, and use the normalizing flow to translate these values back to the space of the original data distribution. A common choice for the latent distribution is the Normal distribution, hence the name *normalizing* flow.

Unlike other deep generative models like Generative Adversarial Networks (GANs; I. J. Goodfellow et al. 117) and Variational Autoencoders (VAEs; D. P. Kingma & M. Welling 161), normalizing flows provide a deterministic mapping between the data and latent spaces. It is this feature of normalizing flows that enables us to provide an exact answer to the question "under the generative model, what is the posterior distribution for redshift given the simulated photometry". This allows us to define a "true" redshift posterior for each galaxy in our simulated catalog, making normalizing flows a very powerful tool for validating photo-z inference.

In this chapter, we introduce PZFlow, a normalizing flow package for Python that is designed to facilitate forward modeling galaxy catalogs with true posteriors for galaxy properties. In addition, we demonstrate PZFlow as a photo-z estimator. With relatively little tuning required by the user, PZFlow can provide a generative model for any tabular data, including continuous and discrete variables, and variables with Euclidean or periodic topology (e.g. the celestial sphere). While calculating posteriors, PZFlow can convolve error distributions and marginalize over missing values.

In Section 3.3 we provide the background on normalizing flows required to understand

PZFlow, which we describe in Section 3.4. In Section 3.5, we demonstrate using PZFlow to simulate a galaxy catalog where each object has a redshift, photometry in the six LSST bands, a true photo-z posterior, a size, and an ellipticity. In Section 3.6, we demonstrate using PZFlow as a density estimator by estimating photo-z's for our simulated catalog. We conclude in Section 3.7.

3.3 Normalizing Flows

Normalizing flows model complex, high-dimensional probability distributions by learning a mapping from the data distribution to a tractable latent distribution[3]. Often the latent distribution is a standard Normal distribution, and so the mapping "normalizes" the data, hence the name "normalizing flow". This mapping allows us to sample and evaluate densities using the latent distribution, rather than the unknown data distribution.

Assume we have a differentiable function f that maps samples x from the data distribution p_X onto samples u from the latent distribution p_U. Using the change of variables formula, we can evaluate the probability density of the data:

$$p_X(x) = p_U(u = f(x)) \, |\det \nabla f(x)|, \tag{3.1}$$

where $\nabla f(x)$ is the Jacobian of f evaluated at x. In words, computing the density $p_X(x)$ is accomplished by mapping x to the latent distribution, calculating its density there, and multiplying by the associated Jacobian determinant, which accounts for how the function f distorts volume elements of the space.

If we further assume that f is invertible, we can sample from the data distribution by applying f^{-1} to samples from the latent distribution[4]:

$$x = f^{-1}(u) \quad \text{where} \quad u \sim p_U. \tag{3.2}$$

---

[3]Some of the machine learning literature defines the mapping in the opposite direction.

[4]Here, $\sim$ means "is drawn from."

Figure 3.1 shows an example of a normalizing flow that transforms the scikit-learn [227] two moons distribution into a uniform distribution. The data points are colored by quadrant to visualize their image under f.

The following sections discuss how to build a normalizing flow to model data with various features. Section 3.3.1 discusses the bijection f and introduces the building blocks from which our bijections will be built; Section 3.3.2 discusses how to choose an appropriate latent distribution for your data; Section 3.3.3 describes how to build a flow that models a conditional distribution; Section 3.3.4 explains how to model data with periodic topology; finally Section 3.3.5 explains how to model data with discrete variables.

3.3.1 Designing a bijection

A bijection is an invertible map between two sets. In a normalizing flow, the bijection maps the data distribution onto the latent distribution for probability calculation, and the inverse of the bijection maps samples from the latent distribution back to the data distribution. The bijection of a normalizing flow must be powerful enough to model complex relationships in data, while remaining invertible and simultaneously possessing an efficiently computable Jacobian determinant. This latter constraint is the primary difficulty in designing a normalizing flow. The most popular strategy for achieving these requirements is to exploit the fact that a composition of bijections is also bijective. By chaining together multiple less-expressive bijections whose Jacobians are efficiently computable, a composite bijections can be constructed that meets our requirements:

$$f = \cdots \circ f_3 \circ f_2 \circ f_1. \tag{3.3}$$

The overall Jacobian determinant can be efficiently calculated using the chain rule.

There is an extensive literature on constructing these sub-bijections (see I. Kobyzev et al. 163 for a review). Some bijections are specialized to be particularly efficient at either density estimation or sampling, but for many science cases, we wish to do both. For this reason, we will focus on Rational-Quadratic Rolling Spline Couplings (RQ-RSCs), bijections which

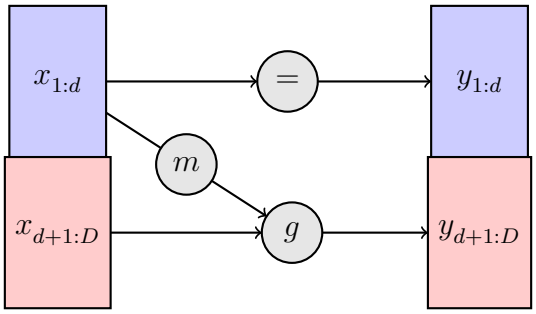

Figure 3.2: Diagram of a coupling layer. The first partition, $x_{1:d}$, is passed through the layer unchanged. The second partition, $x_{d+1:D}$, is transformed by the coupling law g, which is parameterized by the coupling function m applied to the first partition.

achieve state-of-the-art performance, while being efficient with both tasks [95].

Rational-Quadratic Rolling Spline Couplings

RQ-RSCs are bijections that are composed of coupling layers [90, 91]. A coupling layer partitions the data, $x \in \mathbb{R}^D$, into two sets, $x_{1:d}$ and $x_{d+1:D}$. The first set is then used to transform the second set:

$$\begin{aligned} y_{1:d} &= x_{1:d} \\ y_{d+1:D} &= g(x_{d+1:D}; m(x_{1:d})), \end{aligned} \tag{3.4}$$

where $g : \mathbb{R}^{D-d} \times \mathbb{R}^d \to \mathbb{R}^{D-d}$ is an invertible *coupling law*, and m is a *coupling function* defined on $\mathbb{R}^d$. This is illustrated in Figure 3.2. The advantage of this structure is that the Jacobian is triangular,

$$\frac{\partial y}{\partial x} = \begin{pmatrix} I_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial x_{1:d}} & \frac{\partial y_{d+1:D}}{\partial x_{d+1:D}} \end{pmatrix}, \tag{3.5}$$

where I_d is the $d \times d$ identity matrix. In particular, the Jacobian determinant is

$$\det \frac{\partial y}{\partial x} = \det \frac{\partial y_{d+1:D}}{\partial x_{d+1:D}}. \tag{3.6}$$

Furthermore, the inverse can be calculated as

$$\begin{aligned} x_{1:d} &= y_{1:d} \\ x_{d+1:D} &= g^{-1}(y_{d+1:D}; m(x_{1:d})), \end{aligned} \tag{3.7}$$

Notice that neither inverting a coupling layer g, nor calculating the Jacobian determinant, requires inverting or taking derivatives of the coupling function m, which can thus be arbitrarily complex.

The obvious limitation of a coupling layer is that only a subset of the data dimensions are transformed. This is overcome by stacking multiple coupling layers in succession, and switching which variables belong to which partition. In practice, this is achieved by interspersing coupling layers with bijections that shuffle the dimensions of x. These shuffling bijections are trivially inverted and have a Jacobian determinant of one.

In a general coupling layer g, there are a variety of coupling laws m one can use. RQ-RSC's use Rational-Quadratic Neural Spline Coupling [95]. As the name suggests, the coupling law g is a set of rational-quadratic splines. In particular, $g_i : [-B, B] \to [-B, B]$ for each dimension i of $x_{d+1:D}$, where g_i is a piecewise combination of K segments, and each segment is a rational-quadratic function. The positions and derivatives of the knots that parameterize the splines are calculated using the coupling function m, which is a dense neural network applied to $x_{1:d}$.

The result is a bijection that achieves state-of-the-art performance and efficiency for forward modeling and density estimation [163], and are flexible enough to model complex distributions with multiple discontinuities and hundreds of modes. In addition, they are easily adaptable for flows with periodic topology (Section 3.3.4). For more details, see C. Durkan et al. [95].

In this work, we stack Rational-Quadratic Neural Spline Couplings, with Rolling Layers between each – a configuration we name Rational-Quadratic Rolling Spline Couplings (RQ-RSCs). Rolling Layers shift the dimensions of x by one place:

$$\text{Roll} : [x_1, \ldots, x_{D-1}, x_D] \to [x_D, x_1, \ldots, x_{D-1}]. \tag{3.8}$$

By constructing a stack with D coupling layers, RQ-RSCs individually transform each of the D dimensions of x as a function of the other $D-1$ dimensions. This allows the network to learn complex relationships between every subset of the dimensions. In the limit of high spline resolution (i.e. $K \to \infty$), RQ-RSCs can model any differentiable, monotonic function on $[-B, B]^D$ and can thus model arbitrarily complex distributions in this region. In practice, we find very good performance for diverse data sets with $K \approx 16$.

Note you can specify a different value of K for each of the D spline layers in order to individually control the resolution of each dimension. Lowering K typically results in a smoother distribution, while increasing K increases the complexity the normalizing flow can capture, while also increasing computational and memory cost.

Data processing bijections

While RQ-RSCs perform the heavy lifting of mapping the data distribution p_X onto the latent distribution p_U, it is also convenient to define other bijections that perform useful operations such as pre- and post-processing. We name these *data processing bijections*.

For example, RQ-RSCs (and the RQ-NSCs on which they are based) are defined on the domain [-B, B], and thus will not transform samples outside this range. It is therefore useful to define a *Shift Bounds* bijection, which shifts the original range of each dimension to match the domain of the splines. Note this shift must be set at training time, with the assumption that future test data will lie within the same bounds[5]. You can choose a range wider than that covered by the training set if you wish to allow the flow to sample outside the range of the training set

For an example of building an application-specific data processing bijection, see the *Color Transform* bijection defined in Section 3.5.1, which maps galaxy magnitudes to galaxy colors. See section 3.3.5 for data processing bijections that enable modeling of discrete data.

Instead of using these data processing bijections, you can of course manually pre-process

[5]While this sounds quite restrictive, neural networks are typically pretty bad at extrapolating beyond the bounds of the training set anyway.

the data before evaluating densities and post-process samples drawn from the normalizing flow. However, by building pre- and post-processing directly into the bijection, you remove these extra steps from the workflow. This reduces the complexity of working with the normalizing flow and ensures that the flow always "remembers" how to correctly perform these pre- and post-processing steps.

3.3.2 Choosing a latent distribution

In principle, with a sufficiently expressive bijection, the choice of latent distribution does not matter as long as it is a distribution in which you can easily sample and calculate densities. However, in practice, bijections are limited in expressiveness, i.e. they cannot necessarily transform any arbitrary data distribution into any arbitrary latent distribution.

For example, the splines of RQ-RSCs only transform samples in the range [-B, B]. Sampling from a latent distribution with support outside this range will therefore result in strange outliers and incorrect boundary conditions. One can apply a transformation to the latent samples before they are fed into the RQ-RSC to ensure that they lie within the support of the splines, but it is simpler to use a compact latent distribution whose support matches that of the spline layers. A simple choice would be the uniform distribution over [-B, B].

Additionally, as no bijection is perfect, the structure of the latent distribution will not be completely erased in the translation from latent to data distribution. Thus, the latent distribution can be viewed as a prior or inductive bias on samples from the data distribution [151]. It is therefore advantageous to select a latent distribution whose features match some of the structure in the data.

A latent distribution that can achieve both desiderata is the Beta distribution, i.e. $u \sim \text{Beta}(\alpha, \beta)$, where $\alpha, \beta > 0$ are learnable parameters[6]. This distribution is compact, and by varying α and β this distribution can take on a wide variety of shapes with different means, skews, and kurtoses, allowing the inductive bias of the prior to adapt to structure in the

[6]In practice, it is easier to learn $\log \alpha$ and $\log \beta$ to ensure that $\alpha, \beta > 0$.

data during training. However, the Beta distribution is defined on the domain $[0, 1]$, while RQ-RSCs are defined on $[-B, B]$. It is therefore more convenient to use a modified Beta distribution, which we name the *Centered Beta distribution*:

$$\text{CentBeta}(u|\alpha, \beta, B) = 2B \left(\text{Beta}(u|\alpha, \beta) - \frac{1}{2} \right). \tag{3.9}$$

In general, as long as sampling and density evaluation are tractable, one can use any parameterization of the latent distribution that matches some desired structure in the data and learn the distribution parameters during training. We give this generalization the name *latent-adaptive flows* (LAFs; inspired by the Tail Adaptive Flows of P. Jaini et al. 151). Learnable latent distributions can improve training loss, but require more care in training.

Note that while we discussed univariate distributions above, these considerations generalize easily to multiple dimensions. Each of these distributions have multivariate generalizations that can be used when modeling higher-dimensional data. The full multivariate latent distribution can also be assembled by taking the product of multiple univariate distributions[7]. This may even be desired if different dimensions of the data have different structure that you wish to encode in the latent distribution.

3.3.3 Conditional flows

The bijections and latent distributions discussed above can be easily adapted to directly learn conditional probability distributions: you only need to make the replacement $f(x) \to f(x; z)$, where z is a vector of conditions [313]. This is illustrated in Figure 3.3, which is a modification of Figure 3.2 to include the input of conditional variables to the coupling function m. In practice, since m is usually a neural network, this amounts to just appending the conditions z to the inputs of the neural network.

While $p(x|z)$ is technically encoded within $p(x, z)$, which can be learned with a regular normalizing flow, directly modeling $p(x|z)$ with a conditional flow has a few benefits. Training

[7]Note that while the latent variables will be independent, the data variables will still have correlations imprinted by the bijections.

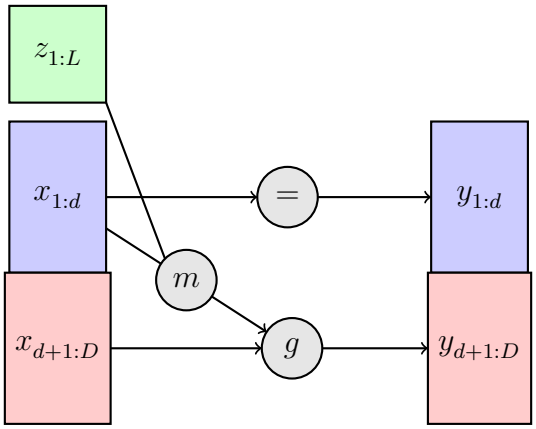

Figure 3.3: Diagram of a *conditional* coupling layer. The first partition, $x_{1:d}$, is passed through the layer unchanged. The second partition, $x_{d+1:D}$, is transformed by the coupling law g, which is parameterized by the coupling function m applied to the first partition *and* the conditional variables $z_{1:L}$. The conditional variables are *never* altered by the flow.

is typically faster, since the latent distribution has a smaller number of dimensions. You can also draw samples of x at fixed values of the conditions z, and you can calculate $p(x|z)$ without having to numerically calculate and divide by $p(z)$, which can be computationally expensive.

3.3.4 Flows with periodic topology

The flows we have considered so far model data that live in $\mathbb{R}^n$. This assumption is insufficient for modeling variables from spaces with non-Euclidean topology, e.g. positions on the sky. While progress has been made on building flows for general topologies (e.g. M. C. Gemici et al. 113 and L. Falorsi et al. 101), we will focus on building flows on the sphere, S^2, as this is the case most relevant in astronomy. We will see that by carefully choosing the latent space, we can construct flows with periodic topology with minimal additional effort [238].

Positions on the sphere are specified by two angles[8], θ and ϕ, the latter of which is

[8]We use the convention where θ and ϕ are the zenith and azimuthal angles, respectively.

periodic. By mapping θ to $\cos\theta$, we map the sphere to a cylinder[9]: $S^2 \to [-1, 1] \times S^1$ (i.e. the Cartesian product of an interval and a circle). In other words, we can transform $\cos\theta$ with a Euclidean flow, as long as we ensure that the flow bounds samples to the range $[-1, 1]$. However, the S^1 piece, ϕ, has a periodic topology and must be handled more carefully.

First, we will address transformations of $\cos\theta$. The only constraint we must impose is that samples of $\cos\theta$ must lie in the range $[-1, 1]$. Fortunately, RQ-RSCs are bounded, mapping a range in u to the same range in x. Thus, if we pick a latent distribution with compact support in $[-1, 1]$, samples of $\cos\theta$ are guaranteed to lie in the same range, as long as we set the range of the RQ-RSC $B = 1$.

Next we will address transformations of ϕ. For f to be a differentiable bijection on the circle, S^1, it is sufficient that f obey the following constraints:

$$f(0) = 0 \tag{3.10}$$

$$f(2\pi) = 2\pi \tag{3.11}$$

$$\nabla f(0) = \nabla f(2\pi) \tag{3.12}$$

$$\nabla f(\phi) > 0. \tag{3.13}$$

The first two constraints ensure continuity of f by designating $\phi = 0$ as a fixed point, and the third constraint ensures continuity of ∇f at that fixed point. While the designation of $\phi = 0$ as a fixed point is an unnecessary restriction on f, any differentiable bijection on the circle has at least one fixed point up to a phase change, and so this restriction does not actually restrict the expressiveness of f. The fourth restriction ensures monotonicity, which guarantees invertibility.

If we make the phase change $\phi \to \phi - \pi$ so that our angles $\phi \in [-\pi, \pi]$, a RQ-NSC with $B = \pi$ automatically fulfills all four constraints. In fact, regular RQ-NSC's impose the further condition

$$\nabla f(-\pi) = \nabla f(\pi) = 1 \tag{3.14}$$

[9]This map can be explicitly constructed via an embedding in $\mathbb{R}^3$. Technically, the map is not defined for $\theta \in \{0, \pi\}$, however as this set has zero measure, it can be safely ignored.

to match an identity transform for inputs outside of the range $[-\pi, \pi]$. By choosing a latent distribution with compact support in the range $[-\pi, \pi]$, we ensure that no samples will lie outside the range of the splines, and so we can relax the boundary condition of Equation 3.14 in favor of the boundary condition in Equation 3.12. Spline transforms with this relaxed boundary condition are named *Circular Splines* by D. J. Rezende et al. [238].

The circular spline construction above is easily generalized to n-spheres and n-tori: $S^n \rightarrow [-1, 1]^{n-1} \times S^1$ and $T^n \rightarrow (S^1)^n$ (see D. J. Rezende et al. 238 for more details). We can model the joint distribution of periodic and non-periodic variables with RQ-RSCs simply by choosing appropriate bounds B for each dimension, and by swapping boundary condition 3.14 for condition 3.12 for any periodic dimensions.

3.3.5 Modeling discrete variables

In addition to the continuous variables described above, normalizing flows can also be used to model discrete variables. This can be achieved by "dequantizing" the discrete dimensions of the data, which can then be mapped onto continuous latent distributions using regular continuous bijections. Dequantization consists of adding some kind of continuous noise to the discrete dimensions, transforming them into continuous dimensions. When sampling from the flow, you simply do the opposite, and "quantize" the discrete dimensions after applying all of the regular bijections, mapping the noisy, continuous variables onto their discrete counterparts.

A common method for dequantization is uniform dequantization, in which random uniform noise in the range (0, 1) is added to the discrete dimensions. The corresponding quantization applied while sampling from the flow consists of applying the floor function to the dequantized dimensions, mapping these samples onto the nearest integer less than the sampled value. More sophisticated dequantization schemes use variational inference or even another normalizing flow to determine the noise distributions, which improves results by smoothing the discontinuities between neighboring discrete values. See J. Ho et al. [134] E. Hoogeboom et al. [136] for more details.

While the dequantizers are not technically bijections, they can be treated as data processing bijections and be chained together with the other bijections in your normalizing flow.

3.4 PZFlow

PZFlow is a Python package for building normalizing flows, with a focus on features useful for forward modeling and density estimation for tabular data. Data is handled in Pandas DataFrames [310], while the normalizing flows are implemented in Jax [34], which allows for efficient, parallelizable, GPU-enabled calculations for very large data sets. The code is easily installable from the Python Package Index[10] (PyPI) and is hosted on Github[11]. The documentation[12] includes tutorial notebooks demonstrating the features mentioned in this chapter on different example problems.

The rest of this chapter will focus on forward modeling a photometric galaxy catalog, and photo-z inference. Section 3.5 uses PZFlow to forward model a galaxy catalog, including photometry, spectroscopic redshifts (spec-z's), true photo-z posteriors, ellipticities, and sizes. Section 3.6 uses PZFlow for photo-z estimation, demonstrating the power of PZFlow as a density estimator, including numerous useful features for photo-z estimation.

In addition to the examples in this chapter, PZFlow has already being used in various other DESC projects:

- A. I. Malz et al. [195] used PZFlow to build a metric for observing strategy optimization based on information theory;

- N. Stylianou et al. [286] used PZFlow to forward model galaxy data with true redshift posteriors in order to evaluate the impact of survey incompleteness and spec-z errors on photo-z estimation;

[10] https://pypi.org/project/pzflow/

[11] https://github.com/jfcrenshaw/pzflow

[12] https://jfcrenshaw.github.io/pzflow/

- M. Lokken et al. [184] used PZFlow to smooth high-redshift artifacts in simulations of host galaxies for supernovae and other transients;

- I. Moskowitz et al. [210] used PZFlow to explore data augmentation for photo-z spectroscopic training sets.

3.5 Forward Modeling a Galaxy Catalog

In this section, we use PZFlow to forward model a photometric galaxy catalog for the Vera Rubin Observatory's Legacy Survey of Space and Time (LSST; Ž. Ivezić et al. 146). The advantage of using a catalog generated from a normalizing flow is that we have direct access to the probability distribution from which the data is drawn, enabling us to calculate true values for derived statistical products, such as the true photo-z redshift posterior for each galaxy.

In Section 3.5.1 we construct a normalizing flow to model the galaxy redshifts and photometry and generate a new simulated catalog. In Section 3.5.2, we calculate true redshift posteriors for the new catalog. In Section 3.5.3 we build a conditional flow to add additional galaxy properties to the catalog.

3.5.1 Forward modeling redshifts and photometry

To create a generative model of galaxy redshifts and photometry, we use the true redshifts and *ugrizy* magnitudes from the CosmoDC2 simulation [165, 188] of the LSST Dark Energy Science Collaboration (DESC). We add photometric errors to the true *ugrizy* magnitudes using the 10-year-depth LSST extended-source error model of our PhotErr package (see Appendix 3.7), and selected galaxies with a signal-to-noise ratio (SNR) greater than 5 in the i band. Of these, we randomly selected 10^6 galaxies and split them into training and test sets consisting of 80% and 20% of the galaxies, respectively. We then train a normalizing flow to learn the distribution $p(z, \hat{\mathbf{m}})$, where z is the true redshift, and $\hat{\mathbf{m}}$ is the vector of noisy magnitudes in the LSST bands.

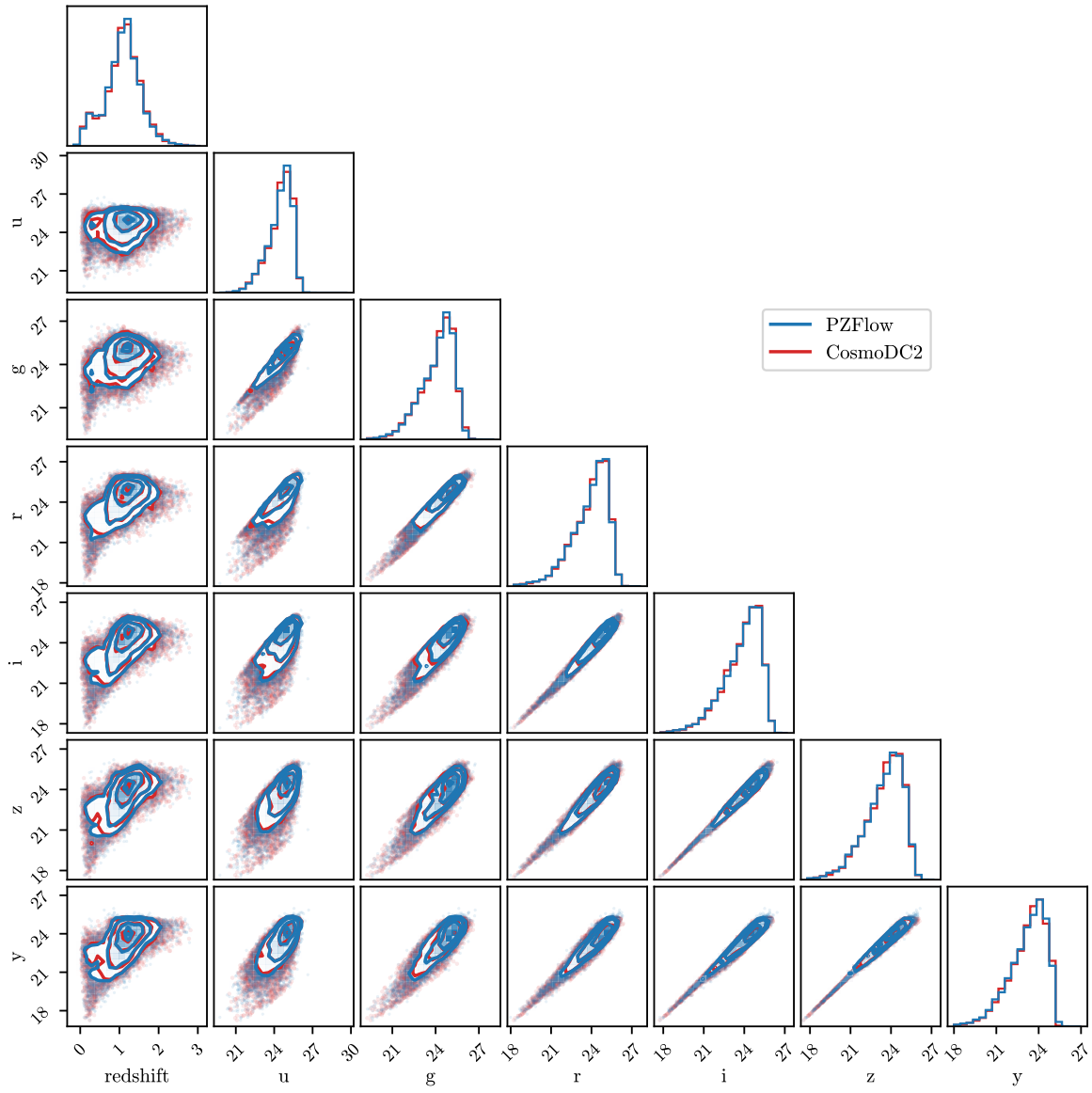

Figure 3.4: Distribution of true redshifts and noisy photometry from the CosmoDC2 test set, compared to a sample drawn from the distribution learned by PZFlow. The close overlap of every pair-wise distribution demonstrates that PZFlow has learned the distribution in CosmoDC2 with high fidelity.

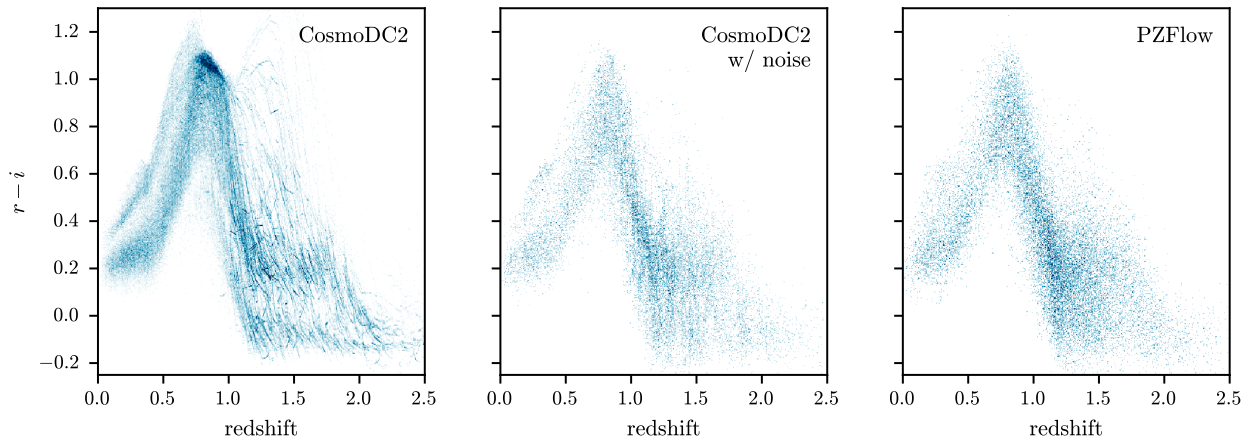

Figure 3.5: Comparison of the $r-i$ vs redshift distribution for galaxy samples from CosmoDC2 without photometric noise (left), CosmoDC2 with photometric noise (middle), and from the normalizing flow (right). High-redshift galaxies in CosmoDC2 lie along discrete tracks in redshift-color space. Adding photometric noise somewhat smooths but does not totally remove these tracks. PZFlow produces a catalog with a smooth redshift-color distribution.

For the latent distribution we use a 7 dimensional Uniform distribution over the range $[-5, 5]$[13]. To map the data onto the latent distribution, we use the following bijection:

$$f = \text{RQ-RSC} \circ \text{Shift Bounds} \circ \text{Color Transform}. \tag{3.15}$$

We will explain each layer of the bijection in the order they are applied to the input data.

The first layer of the bijection is the Color Transform, a data processing bijection designed specifically for this task. The Color Transform converts galaxy magnitudes to colors, but keeps the i band magnitude as a proxy for the apparent luminosity:

$$\begin{aligned}\text{Color Transform} : (\text{redshift}, u, g, r, i, z, y) \rightarrow \\ (\text{redshift}, i, u-g, g-r, r-i, i-z, z-y).\end{aligned} \tag{3.16}$$

[13]The choice of 5 was arbitrary. Any other positive value would work just as well.

This layer is useful as galaxy redshifts correlate more directly with galaxy colors than galaxy magnitudes.

The next layer, Shift Bounds, is the data processing bijection defined in Section 3.3.1, which maps the range of the data onto the support of the RQ-RSC. Note that since Shift Bounds is on the "other side" of the Color Transform, we need to map the ranges of the colors $u - g$, $g - r$, etc. onto the support of the splines, instead of the original magnitudes.

The final layer is an RQ-RSC, described in detail in Section 3.3.1. This layer performs the heavy lifting of transforming the data distribution into the uniform latent distribution. We use $D = 7$ layers to transform all 7 dimensions of our data, and set $B = 5$ to match the support of the latent distribution. We use the coupling function (a feedforward neural network with two hidden layers of 128 neurons) described in C. Durkan et al. [95]. We use $K = 16$ spline knots. This number was chosen to be large enough to capture the complexity of the data, but small enough so that the flow smooths over the discrete tracks in Figure 3.5 (more on this below).

After training the flow (see Appendix 3.7), we assess the results by drawing 10^4 galaxies from the trained flow, and plotting their distribution against 10^4 galaxies from the test set (Figure 3.4). We see the normalizing flow has done an excellent job of reproducing the distribution of galaxies in CosmoDC2, without any unusual artifacts or outliers. In addition, Figure 3.5 compares the distribution of galaxy $r - i$ vs redshift. High-redshift galaxies in the CosmoDC2 simulation lie on discrete tracks in this space due to the way galaxies were assigned to a discrete set of SED templates during simulation. These tracks are easily visible in the left panel. Adding photometric errors somewhat smooths the distribution, but close inspection reveals there is still granularity in the distribution of high-redshift galaxies. The right panel shows the distribution produced by PZFlow, which has a smooth color distribution, even at high redshift. We note that these results were obtained without any extensive hyperparameter search, and that very similar (slightly worse results) are obtained without the `ColorTransform` bijection, demonstrating the flexibility of the method to adapt to unseen data sets.

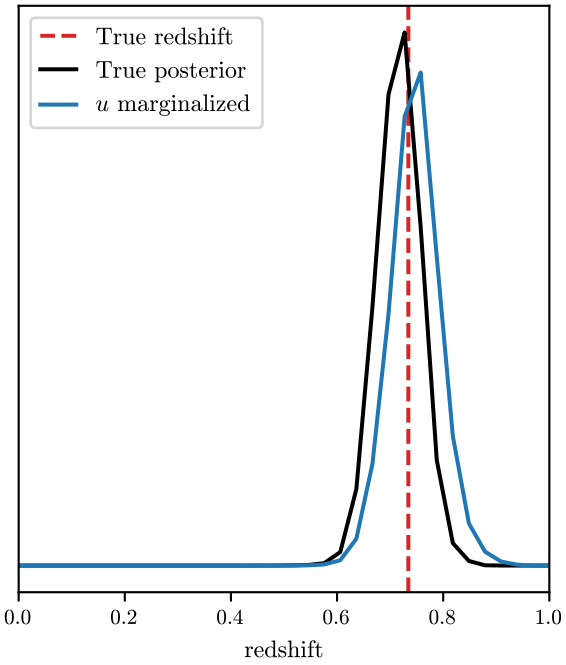

Figure 3.6: Example of redshift posteriors for a galaxy simulated with PZFlow. The true redshift of the galaxy is marked by the vertical dashed red line, and the true redshift posterior for the galaxy is drawn in black. Calculating the posterior while marginalizing over the u band magnitude yields the posterior in blue. Note that the u band marginalization is only approximate, but increasing the resolution of the grid of u band values causes the resultant posterior to converge.

With this normalizing flow, we have an efficient, probabilistic CosmoDC2 emulator that models a smooth color-redshift distribution up to redshift 3. We generate a catalog by sampling 10^4 galaxies from the flow, each with noisy photometry and a true redshift. Importantly, since we have access to the probability distribution from which the galaxies were generated, we can calculate true redshift posteriors for each galaxy. This is the subject of the next section.

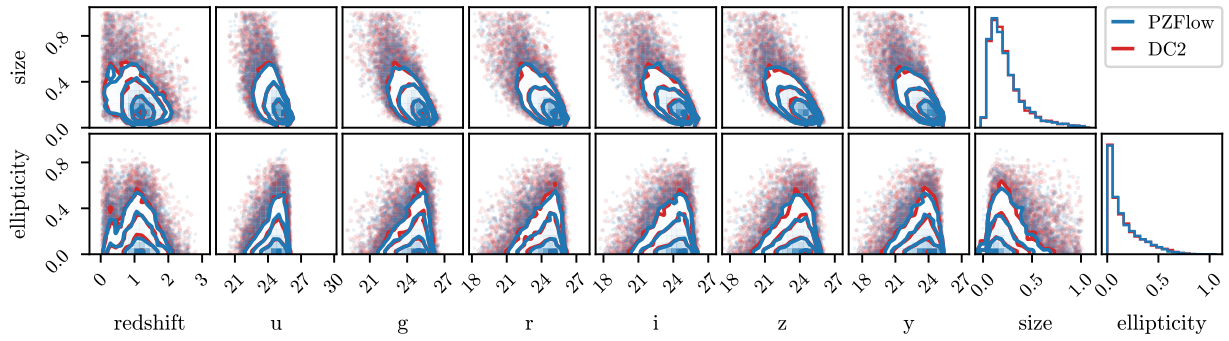

Figure 3.7: Conditional distributions of the ellipticity and size of the galaxies in the CosmoDC2 test set compared to the distribution learned by PZFlow. The close overlap of every pair-wise distribution demonstrates that PZFlow has learned the distribution in CosmoDC2 with high fidelity.

3.5.2 Calculating true posteriors

Since we have direct access to the probability distribution from which the photometry and redshifts are drawn, using Equation 3.1, we can analytically calculate the true redshift posterior for each galaxy: $p(z|\hat{\mathbf{m}})$ where $\hat{\mathbf{m}}$ is the vector of noisy galaxy magnitudes. We note this is not an estimate, like would be returned by a photo-z estimator, but rather the simulated truth, obtained from the model that generated the photometry and redshifts in the first place. Of course the resolution of the posterior is limited to the choice of redshift grid.

When calculating these posteriors, for each galaxy we can also marginalize over the magnitudes in any missing bands. Imagine, for a galaxy, that we partition the vector of magnitudes $\hat{\mathbf{m}}$ into an observed set $\hat{\mathbf{m}}_0$ and a missing set $\hat{\mathbf{m}}_x$. We can marginalize over the missing magnitudes when calculating the posterior

$$p(z|\hat{\mathbf{m}}_0) = \frac{1}{p(\hat{\mathbf{m}}_0)} \int p(z, \hat{\mathbf{m}}_0, \hat{\mathbf{m}}_x) d\hat{\mathbf{m}}_x, \tag{3.17}$$

which can be calculated by evaluating $p(z, \hat{\mathbf{m}})$ on a grid of z and possible values of $\hat{\mathbf{m}}_x$, summing over $\hat{\mathbf{m}}_x$ to yield $p(z, \hat{\mathbf{m}}_0)$, and normalizing with respect to redshift to yield $p(z|\hat{\mathbf{m}}_0)$.

PZFlow possess a flexible method for performing this marginalization: the grid for each band in $\hat{\mathbf{m}}_x$ can be a function of other galaxy properties (e.g. the observed magnitudes, $\hat{\mathbf{m}}_0$).

You may wish to marginalize over all values of $\hat{\mathbf{m}}_x$ if the galaxy was not observed in those bands. This may occur, for example, when simulating a joint Euclid-LSST catalog [263], as not all galaxies will have photometry from both. You may also wish to marginalize over all values beyond the limiting magnitudes to simulate a galaxy that was observed but not detected in the corresponding bands. This might occur, for example, in the low wavelength bands of Lyman-dropout galaxies observed by LSST.

Note this marginalization is only approximate, and therefore weakens our ability to refer to these as "true" redshift posteriors. However, increasing the resolution of the grid used for u band marginalization causes the resultant posteriors to converge. Thus, we believe it is still appropriate to treat these marginalized posteriors as the truth for the purpose of photo-z validation.

Redshift posteriors for an example galaxy are displayed in Figure 3.6. The black posterior is calculated using the full set of galaxy magnitudes. The true redshift, marked by the vertical red line, nearly coincides with the mode of this posterior. The blue posterior has been calculated while marginalizing over the u band. Throwing away the information in the u band slightly broadens the posterior and shifts it toward higher redshifts.

Calculating these posteriors enables direct comparison of true redshift posteriors with the redshift posteriors estimated by photo-z estimators. This is important, as modern cosmology analyses are beginning to increasingly rely on full redshift posteriors [196, 218]. S. J. Schmidt et al. [268] showed that popular metrics for evaluating photo-z estimators using ensembles of photo-z posteriors can be misleading, and are not well suited to the needs of precision cosmology. PZFlow catalogs with true redshift posteriors provide a path forward by enabling the evaluation of photo-z estimators on a per-posterior basis.

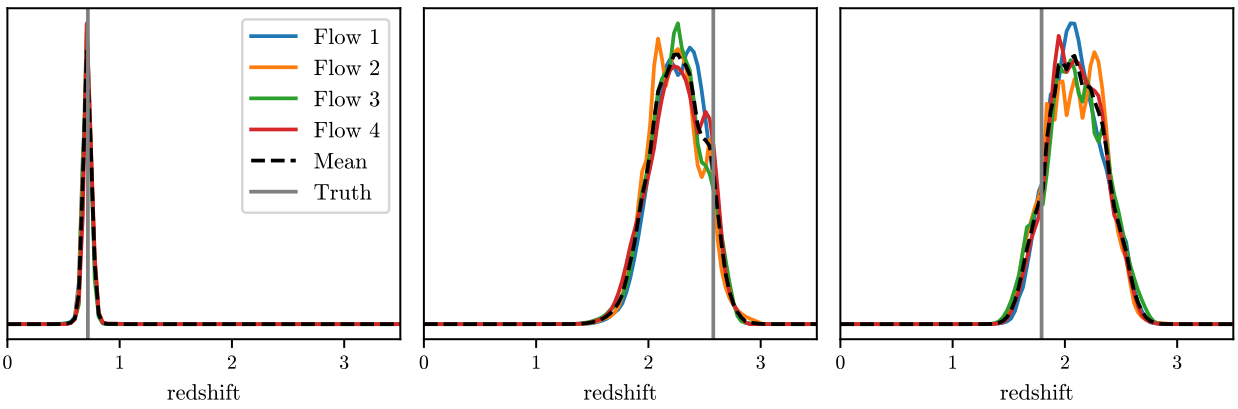

Figure 3.8: The ensemble of posteriors for three example galaxies. Flows 1-4 label the individual posteriors produced by each of the flows that make up the ensemble. The dashed black line is the mean of these individual posteriors and is the value used by the ensemble. The vertical gray line labeled "Truth" denotes the true redshift of the galaxy. Averaging the posteriors from each flow in the ensemble approximately marginalizes over the neural network parameters, and smooths over the small-scale variations found in the posterior from any individual flow. The first panel is a galaxy with a narrow and unimodal redshift posterior, while the next two panels demonstrate broad, multimodal posteriors, which is typical of galaxies in the range $1.5 < z < 2.6$.

3.5.3 Additional properties with conditional flows

In addition to the galaxy magnitudes and redshifts modeled above, we wish to include other galaxy properties in the catalog, such as galaxy size and ellipticity. In principle, we could have included these variables in the original normalizing flow. However, we did not want the true redshift posteriors to be conditioned on these variables, as most photo-z estimators only use galaxy photometry. Therefore, we will build a second flow that models these additional values conditioned on the galaxy redshift and magnitudes. In other words, we are modeling

the full joint distribution via the decomposition

$$p(z, \hat{\mathbf{m}}, s, e) = p(z, \hat{\mathbf{m}}) \cdot p(s, e | z, \hat{\mathbf{m}}), \tag{3.18}$$

where s is the size (the half-light radius in arcseconds) and e is the ellipticity. The first distribution on the right hand side of Equation 3.18 is modeled by our original flow, and the second distribution will be modeled using the new conditional flow. While we have only chosen to model these additional two properties, any other values you desire can be similarly modeled.

For the latent distribution, we again use a Uniform distribution over the range $[-5, 5]$. For the bijection, we use

$$f = \text{RQ-RSC} \circ \text{Shift Bounds}. \tag{3.19}$$

The RQ-RSC acts on the two dimensional space of size and ellipticity, but also takes the galaxy redshift and magnitudes as inputs (see the conditional variables in green in Figure 3.3). The redshifts and magnitudes are transformed to have zero mean and unit variance before being input to the neural network[14] that parameterizes the splines. Aside from the change in inputs, the RQ-RSC has the same settings as listed for the previous normalizing flow.

After training the flow (see Appendix 3.7), we sample a size and ellipticity for each galaxy in the PZFlow catalog created in the previous section (conditioned on the redshift and noisy magnitudes), and plot the distribution of these features against the distribution in the test set (Figure 3.7). Once again, we see the normalizing flow does a good job of emulating the CosmoDC2 galaxy distribution. We note that, if desired, the TARP test of P. Lemos et al. [173] can provide a quantitative test of the fidelity of the conditional flow.

The final simulated catalog consists of 10^4 galaxies, each with a redshift, noisy *ugrizy* magnitudes, a true photo-z posterior, a size, and an ellipticity. We use the magnitudes, size,

[14]These variables are standard scaled instead of mapped onto the domain [-5, 5], because the neural network that parameterizes the splines has no limit on inputs, unlike the splines themselves, which are limited to the range [-5, 5].

and ellipticity to estimate the photometric errors using the 10-year-depth LSST extended-source error model of PhotErr. This small catalog was generated for visualization purposes, but the normalizing flows can be used to generate catalogs of arbitrarily large size. In particular, in only a few minutes, one can generate new catalogs or augment existing catalogs with millions of galaxies. This is substantially faster than re-running large scale simulations like CosmoDC2.

3.6 Photometric Redshift Estimation

In addition to forward modeling, normalizing flows are powerful and flexible models for density estimation. This makes them useful tools for estimating posterior distributions for galaxy properties, conditioned on observed features of the galaxy. In this section, we demonstrate this by applying PZFlow to photo-z estimation for the simulated catalog from the previous section.

3.6.1 Training an Ensemble for photo-z estimation

When forward modeling in Section 3.5, we wanted a realistic model that captured the relevant correlations between galaxy photometry, redshift, shape, and size. However, when estimating redshifts, we do not simply want a realistic model, but rather a model that matches our specific galaxy sample as closely as possible.

When training deep learning models, the huge parameter space contains many different solutions, corresponding to different local minima in the parameter space. In the forward modeling application, we were content with finding a good local minimum, but in this application, we want to marginalize over the different potential models.

A full marginalization over the model parameters would be too computationally expensive, so instead we approximate this marginalization using an ensemble of normalizing flows. In other words, we train multiple normalizing flows under identical conditions, using different random initializations of the model parameters. This allows the optimization algorithm to explore different basins of attraction in the parameter space. In the machine learning

literature, this is known as a Deep Ensemble [168], and is a popular method for approximate bayesian marginalization [107, 311].

We train an ensemble of 4 normalizing flows, each with the same architecture and training schedule as the regular flow described in Section 3.5. With PZFlow, this is as simple as swapping `FlowEnsemble` for `Flow` in the code.

For the training set, we use 100,000 galaxies from the catalog created in Section 3.5. Each galaxy in the training set has a true redshift and observed noisy magnitudes in the *ugrizy* bands, with corresponding photometric errors. To account for the photometric error, at the start of each training epoch, we resample the training set from the photometric error distributions. In other words, each epoch, for each galaxy, we sample

$$\mathbf{m} \sim p(\hat{\mathbf{m}}, \sigma_{\mathbf{m}}), \tag{3.20}$$

where $\hat{\mathbf{m}}$ are the observed magnitudes with photometric errors $\sigma_{\mathbf{m}}$, and $p(\hat{\mathbf{m}}, \sigma_{\mathbf{m}})$ is a Gaussian in flux space. This allows our ensemble of flows to approximate the distribution $p(z, \mathbf{m})$, where $\mathbf{m}$ is the vector of true magnitudes for the galaxy. For more details on training the ensemble, see Appendix 3.7.

3.6.2 Estimating posteriors

After training, we use each flow in the ensemble to estimate the redshift posterior by marginalizing over the photometric errors:

$$p(z|\hat{\mathbf{m}}, \sigma_{\mathbf{m}}) \propto \int p(z, \mathbf{m})\, p(\mathbf{m}|\hat{\mathbf{m}}, \sigma_{\mathbf{m}})\, d\mathbf{m}, \tag{3.21}$$

which is estimated by sampling $\mathbf{m} \sim p(\hat{\mathbf{m}}, \sigma_{\mathbf{m}})$ and averaging $p(z, \mathbf{m})$ over these samples. We then average the $p(z, \mathbf{m})$ from each flow, and normalize with respect to the redshift grid. This provides a redshift posterior for each galaxy.

Posteriors for three galaxies can be seen in Figure 3.8. Each flow produces a PDF which may contain slightly different features in each case. By averaging over the individual posteriors, we select for features that are common between models, while smoothing over

features that are present in only a single model. The first example galaxy in Figure 3.8 is at $z < 1$, and all flows in the ensemble return essentially the same narrow redshift posterior. This is typical for low-redshift galaxies whose photo-z's are relatively well constrained by LSST photometry.

The other two example galaxies, however, are in the $1.5 < z < 2.6$ redshift range, where the Balmer Break (at ~ 4000 Å) has redshifted out of LSST's wavelength coverage, while the Lyman Limit (at 912 Å) has not yet redshifted into LSST's wavelength coverage. As a result, these posteriors are much broader and less well constrained. For these two galaxies, the flows in the ensemble return posteriors with different small-scale variations, and the best-estimate redshift (i.e. the mode of the redshift posterior) varies by as much as 0.5 for each set of posteriors. Marginalizing over the individual posteriors smooths over these variations. We can also treat the ensemble of posteriors as a distribution over possible posteriors, which will allow for more consistent error calibration in cosmological analyses [321].

3.6.3 Photo-z metrics

In this section, we evaluate the performance of PZFlow using common photo-z metrics. Note these metrics are optimistic in the sense that the training set is representative of the test set, which is usually not the case in modern cosmology applications.

The most common metrics for photo-z estimation concern photo-z point estimates, which are a compression of the photo-z posterior to a single redshift estimate (e.g., H. Hildebrandt et al. 132, C. Sánchez et al. 261). We make the common choice of selecting the mode of the posteriors[15]. We compute metrics of the quantity $\Delta z = (z_{\rm phot} - z_{\rm true})/(1 + z_{\rm true})$, where the denominator accounts for naturally greater uncertainties at high redshift.

Figure 3.9 compares the photo-z point estimates to the true redshifts. The point estimates for most galaxies lie along the diagonal, indicating strong performance. There are the common photo-z "wings", indicating redshifts where important spectral features are

[15]The mean redshift is a poor choice, since photo-z posteriors are often multimodal, and so the mean value can lie between two modes at a redshift with very small probability density.

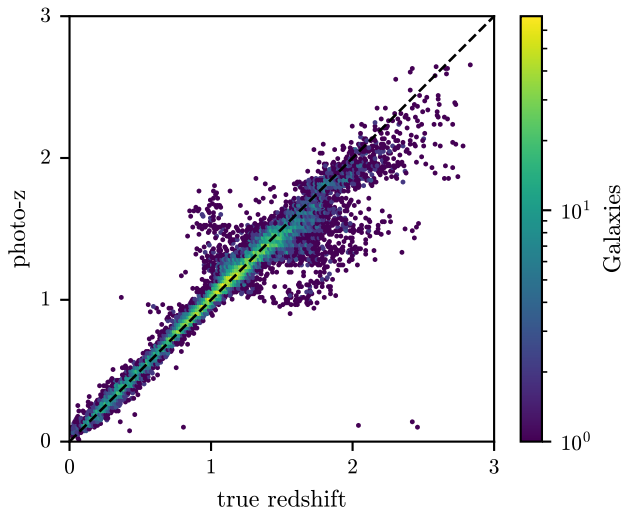

Figure 3.9: Photo-z point estimates (maximum a posteriori) vs true redshift for galaxies in the test set.

transitioning between neighboring photometric bands. This point estimate plot is comparable to other high-performance machine learning photo-z estimators when provided with representative training sets [261].

Figure 3.10 shows the photo-z point estimate metrics from the LSST DESC Science Requirements Document (SRD; The LSST Dark Energy Science Collaboration et al. 291) as a function of true redshift. The *bias* is defined as the median of Δz; the *scatter* is defined as IQR/1.349, where IQR is the interquartile range of Δz; the *outlier fraction* is defined as the fraction of galaxies for which Δz is greater than three times the scatter. The requirements from the SRD are plotted in black to provide a sense of scale.

Like many photo-z estimators, PZFlow performs well to a redshift of approximately 1.5 (the scatter is high at low redshifts due to the relatively small number of low-redshift galaxies in our training set). At higher redshifts, our estimator does not meet the bias and scatter requirements, because there is very little training data in this redshift range. We note however that for many cosmology applications, it is okay for the bias to exceed the required

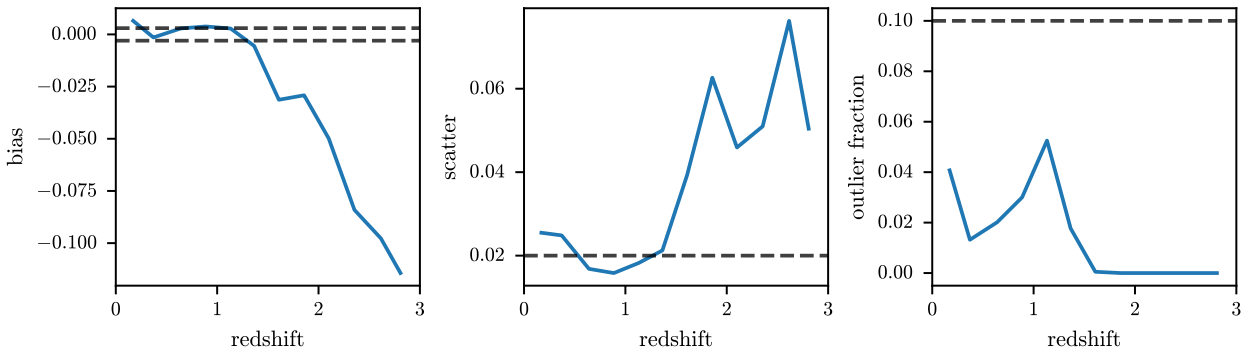

Figure 3.10: The bias, scatter, and outlier fraction of the photo-z point estimates as a function of true galaxy redshift. The dashed black lines represent the limits for LSST cosmology as stated in the LSST DESC SRD [291]. The bias must be between these dashed lines, while the scatter and outlier fraction must be below the dashed lines. You can see that PZFlow meets the bias and scatter requirements up to redshift ~ 1.5 (the scatter is high at low redshifts due to the relative small number of low-redshift galaxies in our training set). PZFlow meets the outlier requirements for all redshifts. We note that individual redshifts do not actually need to meet the bias requirement as long as the bias can be well calibrated via some other source, e.g. galaxy clustering.

limits, as long as the bias can be well determined via some calibration process [220].

Another common metric is the probability integral transform (PIT) (see e.g. S. J. Schmidt et al. 268, B. Dey et al. 87), which is a histogram of the cumulative density function (CDF) of each posterior. I.e., given an estimated posterior $p(z|\hat{\mathbf{m}}, \sigma_{\mathbf{m}})$ and the true redshift z_{true}, the CDF is

$$\text{CDF} = \int_0^{z_{\text{true}}} p(z|\hat{\mathbf{m}}, \sigma_{\mathbf{m}})\, dz. \tag{3.22}$$

For perfectly calibrated posteriors, the CDF distribution (the PIT histogram) is uniform between 0 and 1. This is because, for example, if the photo-z posteriors produced by an estimator are well calibrated you would expect the true redshifts of the galaxies to fall within the 50% confidence intervals 50% of the time.

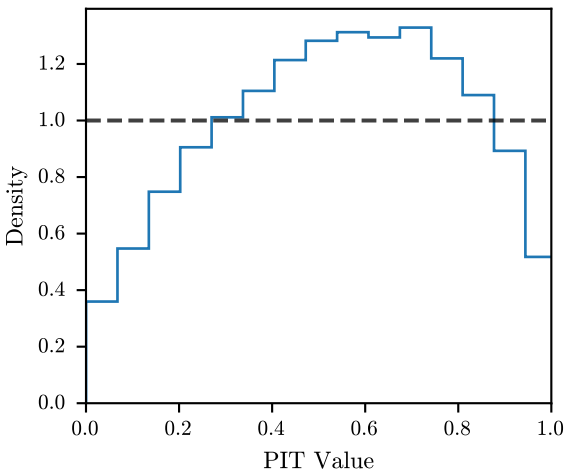

Figure 3.11: The probability integral transform (PIT) histogram for PZFlow photo-z posteriors. The PIT characterizes the calibration of the estimated posteriors, with the horizontal black line indicating perfect calibration.

The PIT histogram for our estimator is shown in Figure 3.11. Ideally, this histogram would be uniform and match the dashed horizontal line, which represents perfect calibration. The fact that the histogram bulges at the center indicates our estimator is too conservative – i.e. the posteriors it produces are too broad. This can be explained by the fact that normalizing flows exhibit mode covering behavior (the opposite of the mode collapse seen in GANs; T. Salimans et al. 259). In other words, because normalizing flows are trained by maximizing the likelihood of the training samples, they receive very high penalties for missing any modes in the data. As a result, they tend to conservatively spread out their density, in order to avoid missing any modes. This results in overly conservative posterior predictions.

The low values at the edges of the PIT histogram indicate the relative rarity of catastrophic outliers, which is also reflected in the far right panel of Figure 3.10, where you can see our estimator meets the requirement on the outlier fraction at all redshifts. There is also a slight rightward tilt. This indicates a small negative bias, which reflects the intrinsic

prior towards smaller redshifts, as this is where the majority of galaxies in the training set lie. This negative bias is visible for high-redshift galaxies in the far left panel of Figure 3.10. Calibrating these posteriors, either via altering the training loss or post-processing the posteriors, is beyond the scope of this chapter. However this calibration could be achieved, for example, using the methods of B. Dey et al. [87].

The previous metrics analyze photo-z performance for point estimates, which are insufficient for modern cosmology [218], and for ensembles of posteriors, which is often misleading and not a good indicator of performance for science applications [268]. The methods introduced in this chapter enable the creation of galaxy catalogs for which each galaxy has a true redshift posterior, which will enable more comprehensive evaluation of photo-z estimators. Full evaluation of photo-z estimators on a posterior-by-posterior basis is a major goal of the LSST DESC, and will be the focus of forthcoming work.

3.7 Conclusion

In this chapter we introduced PZFlow, a Python package for probabilistic forward modeling of galaxy catalogs, and demonstrated how it will be used to assist the photo-z calibration efforts of the LSST DESC. In particular, galaxies generated from a PZFlow model have a natural notion of a true photo-z posterior, to which the redshift posteriors estimated by photo-z algorithms can be directly compared. This enables a more comprehensive evaluation of the posteriors produced by photo-z estimators that we expect will avoid the traps of ensemble-only metrics that were identified by S. J. Schmidt et al. [268]. Validating the full posteriors produced by photo-z estimators is vital for enabling unbiased cosmology inference with next generation surveys like the LSST [218]. Synthetic catalogs from PZFlow, together with new metrics of posterior calibration (e.g. the tests of local conditional calibration of B. Dey et al. 86, B. Dey et al. 87) will be used in future data challenges to optimize and to quantify the error rate and biases of the DESC photo-z estimation pipeline.

In addition to forward modeling, PZFlow is a powerful tool for density estimation applied to tabular data. We demonstrated this by applying PZFlow to the task of photo-z

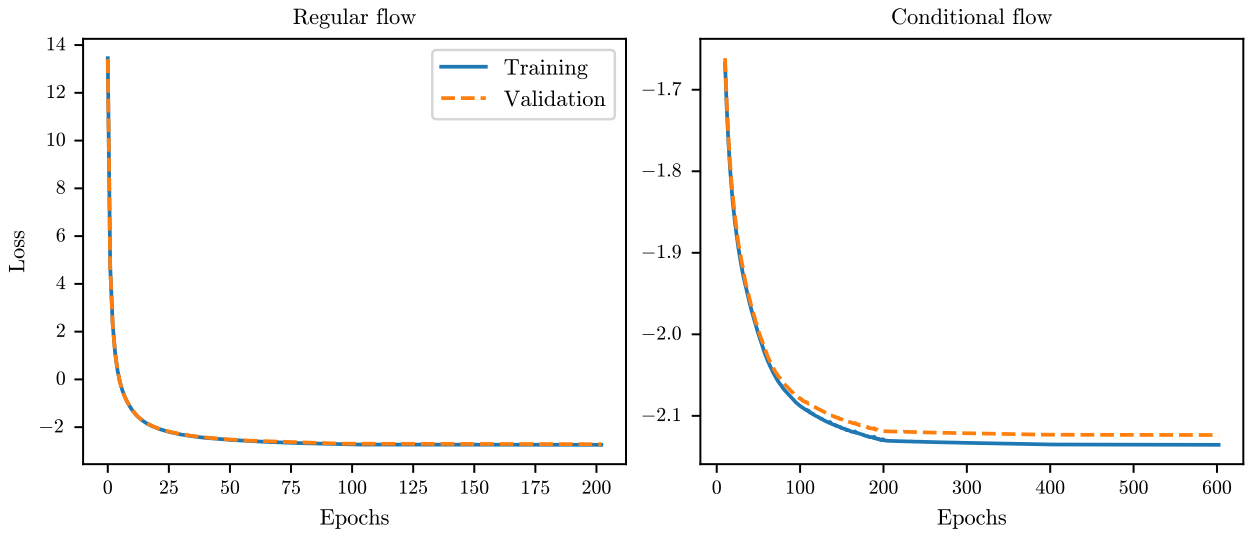

Figure 3.12: Training losses for the normalizing flows used to simulate the galaxy catalog. Left: losses for the regular flow. Right: losses for the conditional flow.

estimation. PZFlow achieves high accuracy with very little fine tuning and very few modeling assumptions. However, as PZFlow is trained via likelihood maximization of the training set, it exhibits mode-covering behavior — i.e., in order to not miss any modes in the data, PZFlow tends to be conservative and produce overly broad posteriors. Increasing the amount of training data will likely alleviate these issues, but tradeoffs of this variety are inherent in any choice of machine learning model [268].

While we have developed PZFlow to address the calibration needs of DESC photo-z validation, and have focused on those applications in this chapter, we emphasize that PZFlow is a powerful and flexible tool for statistical modeling of any tabular data.

Appendix 3.a: Flow training details

In this section we list some technical details of training the normalizing flows. Every flow is trained via minimizing the negative log-likelihood

$$\mathcal{L} = -\,\mathbb{E}[\log p(x)\,], \tag{3.23}$$

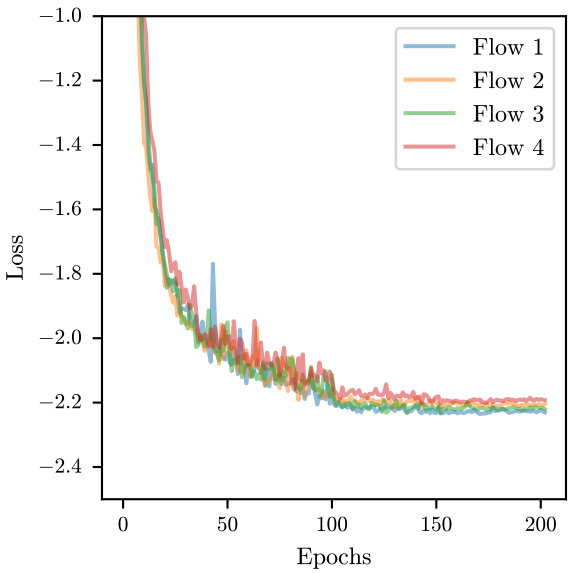

Figure 3.13: Training losses for the four flows in the flow ensemble. We have zoomed in to the bottom of the loss curve so you can see that each of the flows converges to a slightly different minimum loss.

where the expectation is performed over galaxies in the training set and $p(x)$ is defined in Equation 3.1.

For the main flow in Section 3.5, we trained for 200 epochs. We used the Adam optimizer [160], starting with a learning rate of 10^{-5}. We decreased the learning rate by a factor of 10 after the 100th and 150th epochs. Training took 7 minutes on a Tesla P100 12GB GPU. The training loss for this flow is in the left panel of Figure 3.12.

For the conditional flow in 3.5, we trained for 600 epochs. Again, we used Adam with an initial learning rate of 10^{-5}. We decreased the learning rate by a factor of 10 every 200 epochs. The training loss for this flow is in the right panel of Figure 3.12.

For each of the flows that make up the flow ensemble in Section 3.6, we trained for 200 epochs using the Adam optimizer. We started each with a learning rate of 10^{-4}, which we decreased by a factor of 10 after the 100th and 150th epochs. The training loss for the

ensemble is in Figure 3.13. Each flow achieved nearly the same training loss.

Appendix 3.b: LSST Error Model

We estimate photometric errors for LSST using a generalization of the error model from Ž. Ivezić et al. [146]. To derive the error model, we start with the noise-to-signal ratio (NSR) for an object with photon count C and background noise N_0 (which depends on seeing, read-out noise, etc.):

$$\mathrm{NSR}^2 = \frac{N_0^2 + C}{C^2}. \tag{3.24}$$

If we define $C = C_5$ when $\mathrm{NSR} = 1/5$, then we can solve for N_0 and write

$$\mathrm{NSR}^2 = \frac{1}{C_5}\left(\frac{C_5}{C}\right) + \left[\left(\frac{1}{5}\right)^2 - \frac{1}{C_5}\right]\left(\frac{C_5}{C}\right)^2. \tag{3.25}$$

Defining $x = C_5/C = 10^{(m-m_5)/2.5}$ and $\gamma = 1/5^2 - 1/C_5$, we have

$$\mathrm{NSR}^2 = (0.04 - \gamma)\, x + \gamma\, x^2 \quad (\mathrm{mag}^2), \tag{3.26}$$

which is Equation 5 from Ž. Ivezić et al. [146]. Values for the band-dependent parameter γ can be found in Table 2 of the same paper.

In the high signal-to-noise (SNR) limit, $\mathrm{NSR} \ll 1$, and we can approximate

$$\sigma_{\mathrm{rand}} = 2.5 \log_{10}(1 + \mathrm{NSR}) \approx \mathrm{NSR}. \tag{3.27}$$

This latter approximation is made by Ž. Ivezić et al. [146], and errors are assumed to be Gaussian in magnitude space. In contrast, we use the exact form of Equation 3.27, and model errors as Gaussian in flux space. Note that after the photometric errors are applied, the error is re-calculated from the "observed" flux, and this new error is reported as the estimated photometric error. If the original photometric error were reported, it would provide a deterministic link to the original flux.

We have implemented this error model, along with several other extensions, in the Python package PhotErr, which is available on the Python Package Index[16] (PyPI), and Github[17]. The extensions include different methods for handling non-detections, methods for modeling errors of extended objects (using models from K. Kuijken et al. 166, J. L. van den Busch et al. 304), and error models for the Roman and Euclid space telescopes [120, 263, 279].

[16] https://pypi.org/project/photerr/

[17] https://github.com/jfcrenshaw/photerr

Chapter 4

QUANTIFYING THE IMPACT OF LSST U-BAND SURVEY STRATEGY ON PHOTOMETRIC REDSHIFT ESTIMATION AND THE DETECTION OF LYMAN-BREAK GALAXIES

In 2024 the Rubin Survey Cadence Optimization Committee (SCOC) initiated a study on whether the survey strategy for u-band observations should be altered to better optimized LSST science goals. The following chapter details work that was submitted to the SCOC arguing for increasing the survey time dedicated to u-band observations, which would improve high-redshift photo-z estimation and the detection of redshift $z \sim 3$ LBGs. This work contributed to the SCOC's recommendation to increase the time spent in the u-band by 39%, yielding significant gains for high-redshift cosmology. This chapter also produces estimates for the number density of LSST-detected high-redshift LBGs which are important for the cosmology forecasts in the following chapter.

This chapter is available on the arxiv as J. F. Crenshaw et al. [72] and has been submitted to a special issue on LSST survey strategy in the Astrophysical Journal Supplement series (ApJS).

4.1 Abstract

The Vera C. Rubin Observatory will conduct the Legacy Survey of Space and Time (LSST), promising to discover billions of galaxies out to redshift 7, using six photometric bands ($ugrizy$) spanning the near-ultraviolet to the near-infrared. The exact number of and quality of information about these galaxies will depend on survey depth in these six bands, which in turn depends on the LSST survey strategy: i.e., how often and how long to expose in each band. u-band depth is especially important for photometric redshift (photo-z) estimation

and for detection of high-redshift Lyman-break galaxies (LBGs). In this chapter we use a simulated galaxy catalog and an analytic model for the LBG population to study how recent updates and proposed changes to Rubin's u-band throughput and LSST survey strategy impact photo-z accuracy and LBG detection. We find that proposed variations in u-band strategy have a small impact on photo-z accuracy for $z < 1.5$ galaxies, but the outlier fraction, scatter, and bias for higher redshift galaxies varies by up to 50%, depending on the survey strategy considered. The number of u-band dropout LBGs at $z \sim 3$ is also highly sensitive to the u-band depth, varying by up to 500%, while the number of $griz$-band dropouts is only modestly affected. Under the new u-band strategy recommended by the Rubin Survey Cadence Optimization Committee, we predict u-band dropout number densities of 110 deg^{-2} (3200 deg^{-2}) in year 1 (10) of LSST. We discuss the implications of these results for LSST cosmology.

4.2 Introduction

The Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) will survey approximately 18,000 deg^{-2} of the southern sky to unprecedented depth across such a wide area [146]. This is made possible by an 8.4 m primary mirror and 3.2 Gigapixel camera, yielding an etendue significantly larger than any other existing telescope [189]. Six photometric bands spanning the near-ultraviolet to near-infrared will provide information about the spectral energy distributions of objects detected by LSST, enabling, among other things, photometric redshift (photo-z) estimation for billions of galaxies. The quality of this information, however, depends on the depth in each of these bands, which in turn depends on the survey strategy. Of particular importance for static science is the number and length of visits to each field in the survey footprint.

Over many years, the Rubin Survey Cadence Optimization Committee (SCOC), in collaboration with the Rubin science community, has refined LSST survey strategy to maximize Rubin's science potential, while balancing the needs of a diverse set of science cases (e.g., F. B. Bianco et al. 25, R. L. Jones et al. 154, M. Lochner et al. 182, LSST Science Collabora-

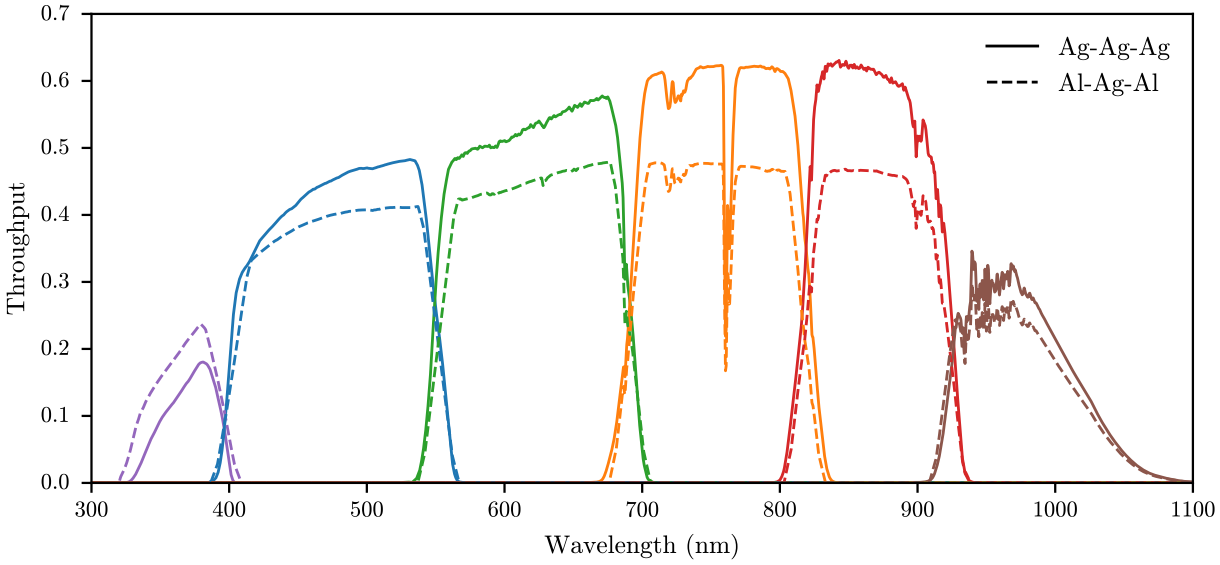

Figure 4.1: Comparison of Rubin Observatory *ugrizy* throughput curves, assuming original Al-Ag-Al and new Ag-Ag-Ag mirror coatings. The transition to all-silver coatings decreased throughput in the *u* band, but increased throughput in all of the *grizy* bands. These curves include contributions from the atmosphere (assuming airmass 1.2), mirror reflectivities, lens and filter throughputs, and detector sensitivity.

tion et al. 190, Rubin Observatory Survey Cadence Optimization Committee 250, 251, 252). This is an ongoing process that will continue throughout the 10 year duration of LSST. Recently, the decision was made to change the coating of Rubin's primary (M1), secondary (M2), and tertiary mirrors (M3)[1] from aluminum-silver-aluminum (Al-Ag-Al) to silver-silver-silver (Ag-Ag-Ag)[2]. This change reduces the throughput in the ultraviolet (i.e., in Rubin's *u* band), while increasing throughput at longer wavelengths (Rubin's *grizy* bands; see Fig. 4.1). Because substantially more survey time is spent observing in *grizy* bands than in the *u* band alone, this results in greater survey efficiency, enabling deeper imaging by the end of LSST.

[1]Rubin's primary and tertiary mirrors are a singular structure, usually referred to as the M1M3.

[2]Note this decision was made before the mirrors were ever coated, so the M1M3 mirror was never actually coated with aluminum.

Under the nominal survey strategy at the time of this change, however, the change in mirror coating reduced the 10-year u-band depth by 0.21 magnitudes.

The u band, spanning approximately 3300 - 4000 Å, is important for a variety of science cases. At low redshifts, the redshifting of the ~ 4000 Å Balmer break through the u band enables accurate photo-z estimation for galaxies at redshifts $z < 0.5$ [156]. Without the u band, photo-z's for these galaxies are highly uncertain, resulting in catastrophic outliers that significantly degrades photo-z estimation up to redshifts $z < 0.6$ [189]. This has implications for nearly every area of LSST cosmology, including lens and source selection for galaxy clustering and lensing analyses [174, 268, 321, 324], galaxy cluster detection [100], supernova cosmology [49, 50, 205]. Achieving precision cosmology therefore places stringent constraints on photo-z performance [218, 220, 291].

At higher redshifts, Lyman-series absorption lines shift into the u band, beginning with the Lyman-alpha transition at rest-frame 1216 Å, which redshifts into the u band at $z \sim 1.7$, and culminating with the Lyman-limit at rest-frame 912 Å, which redshifts into the u band at $z \sim 2.6$. These features, caused by absorption from neutral hydrogen in the intergalactic medium (IGM) and within the galaxies themselves, enable identification and photo-z estimation for high-redshift star-forming galaxies known as Lyman-break galaxies (LBGs; C. C. Steidel et al. 282).

For thirty years, LBGs have been important for studies of galaxy formation and evolution [94, 115], including the ultra-high-redshift LBGs discovered in recent years by JWST, illuminating galaxy formation at cosmic dawn (e.g., S. L. Finkelstein et al. 106, Y. Harikane et al. 130, C. A. Mason et al. 198). LBGs, however, are also increasingly being recognized as cosmologically important tracers of the matter density field at high redshifts. With the width and depth of LSST, using LBGs as cosmological tracers has the potential to provide unprecedented constraints on the growth of structure and evolution of dark energy at redshifts $2 < z < 6$ [206, 312]; competitive constraints on the amplitude of local-type primordial non-Gaussianity [48, 269]; constraints on the sum of neutrino masses, independent of the optical depth of reionization [320]; constraints on the masses of $z > 1$ clusters [301]; and

constraints on the low-redshift matter density field via inverse galaxy-galaxy lensing (IGGL; D. N. Cross & C. Sánchez 73).

The u-band dropout technique, which selects galaxies with a strong u-band flux deficit compared to the g or r band flux, is especially effective at selecting LBGs at redshifts $2 < z < 4$. The success of this technique is strongly dependent on u-band depth. By selecting u-band dropouts in images from the CFHT Large Area u-band Survey (CLAUDS; M. Sawicki et al. 262) and deep grz imaging from Hyper Suprime Cam Subaru Strategic Program (HSC SSP; H. Aihara et al. 9), V. Ruhlmann-Kleider et al. [253] demonstrated it is possible to achieve a spectroscopically-confirmed $2.3 < z < 3.5$ LBG number density of 620 deg^{-2}, from an initial photometric sample of $1,100$ deg^{-2} with $r < 24.2$. The 10-year LSST Wide Fast Deep (WFD) survey, however, will be shallower than the CLAUDS and HSC SSP deep fields. More recently, C. Payerne et al. [226] employed a refined LBG selection method on shallower imaging, simulating the ongoing Ultraviolet Near Infrared Optical Northern Survey (UNIONS, Gwyn et al., in prep.) which will have a depth similar to LSST year 2. This work achieved a confirmed LBG number density of 493 deg^{-2} within $2 < z < 3.5$ from a photometrically-selected sample of $1,100$ deg^{-2} with $r < 24.3$, providing valuable insights into LSST's near-future potential for probing dark energy, growth of structure, and primordial non-Gaussianity.

It is therefore essential to understand how LSST survey strategy impacts u-band depth and the corresponding implications for photo-z estimation and high-redshift cosmology. Due to the increased survey efficiency provided by the Ag-Ag-Ag mirror coatings, it is possible to allocate observing time in each band such that the coadded depths in all bands are deeper than the nominal depths assuming the original Al-Ag-Al mirror coatings. The question, therefore, is how to best balance the depths in each of the six bands to optimize photo-z estimation and LBG detection. In this chapter, we use simulations and simple analytic models to address these questions. Section 4.3 details the simulations and models used in this chapter. In Section 4.4 we describe the metrics used to quantify photo-z performance and LBG detection for different observing strategies, the results of which are presented in

Section 4.5. We discuss the implications of our results for LSST cosmology and conclude in Section 4.6.

We assume Planck Collaboration et al. [232] cosmology throughout. We use AB magnitudes, and refer to apparent magnitudes, N-sigma depths, and cuts in LSST bands as u, u_N, $u_{\rm cut}$, etc. We use lowercase m to refer to apparent magnitudes in arbitrary bands (i.e. any of *ugrizy*), and uppercase M to refer to absolute magnitudes at rest-frame 1500 Å. For each LBG dropout sample, we use "dropout band" to refer to the Rubin band in which the redshifted Lyman-break falls, and "detection band" to refer to the band closest to rest-frame 1500 Å. Specifically, for *ugriz* dropout samples, the dropout bands are *ugriz* and the detection bands are *rizzy*. This notation is summarized in Table 4.1.

Finally, we note this chapter contributes to an ongoing literature studying how survey strategy impacts photo-z estimation, including M. L. Graham et al. [119, 120], Q. Hang et al. [127], M. Lochner et al. [181, 182], A. I. Malz et al. [195], B. R. Scott et al. [273].

4.3 Simulations and IGM + LBG Models

Different survey strategies are simulated using the Rubin Observatory Operations Simulator, described in Section 4.3.1. The impact on photo-z estimation is studied using a simulated galaxy catalog, which is described in Section 4.3.2, including modeling the effects of IGM extinction in the u and g bands. High-redshift LBGs are studied using an analytic model, described in Section 4.3.3, which enables consistent comparison between observing strategies and obviates the need for a simulated high-redshift catalog.

4.3.1 Rubin Operations Simulator

The Rubin Observatory Operations Simulator[3] (OpSim) generates mock observations over the 10-year duration of LSST. These simulations include telescope movement, dithering, variable weather and seeing conditions drawn from a Cerro-Tololo Inter-American Observatory

[3] https://rubin-sim.lsst.io/

Notation	Description
u, g, r, i, z, y	Apparent magnitudes in LSST bandpasses
$u_N, g_N, \ldots$	Nσ limit in LSST bandpasses
$u_{\rm cut}, g_{\rm cut}, \ldots$	Cuts applied to LSST bandpasses
m	Apparent magnitude in an arbitrary band
m_N	Nσ limit in an arbitrary band
$m_{\rm cut}$	Cut applied to an arbitrary band
M	Absolute magnitude at rest-frame 1500 Å
"Dropout band"	Bandpass in which the Lyman-break falls (i.e., containing $(1+z)\,912$ Å)
"Detection band"	Bandpass closest to rest-frame 1500 Å(i.e., closest to $(1+z)\,1500$ Å)

Table 4.1: Summary of notation used in this chapter. All magnitudes are in the AB system; for *ugriz*-dropouts, the detection bands are *rizzy*, respectively.

Strategy u time	Relative Δu	Δg	Δr	Δi	Δz	Δy	
1.0x, 30s u	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1.0x, 38s u	1.27	0.18	-0.01	-0.01	-0.01	-0.01	-0.00
1.0x, 45s u	1.50	0.31	-0.01	-0.02	-0.02	-0.02	-0.01
1.0x, 60s u	2.00	0.52	-0.04	-0.04	-0.03	-0.04	-0.02
1.1x, 30s u	1.10	0.04	-0.01	-0.00	-0.01	0.00	-0.00
1.1x, 38s u	1.39	0.24	-0.02	-0.01	-0.01	-0.01	-0.00
1.1x, 45s u	1.65	0.36	-0.02	-0.02	-0.02	-0.02	-0.01
1.1x, 60s u	2.20	0.55	-0.04	-0.04	-0.04	-0.04	-0.02
1.2x, 30s u	1.20	0.09	-0.01	-0.01	-0.01	-0.01	-0.01
1.2x, 38s u	1.52	0.29	-0.03	-0.02	-0.02	-0.02	-0.01
1.2x, 45s u	1.80	0.41	-0.03	-0.02	-0.03	-0.03	-0.02
1.2x, 60s u	2.40	0.61	-0.05	-0.04	-0.05	-0.05	-0.04
1.5x, 30s u	1.50	0.22	-0.01	-0.02	-0.04	-0.01	-0.01
1.5x, 38s u	1.90	0.40	-0.04	-0.04	-0.04	-0.03	-0.02
1.5x, 45s u	2.25	0.53	-0.05	-0.04	-0.05	-0.04	-0.03

Table 4.2: u-band strategy variations to baseline v3.4. The second columns quantifies the time spent observing in the u band, relative to the baseline v3.4 observing strategy. The right six columns display changes in year 10 *ugrizy* median 5σ depths for the WFD survey. For comparison, the median 5σ depths for the [1.0x, 30s u] strategy are 25.15, 26.65, 26.84, 26.40, 25.73, 24.78 for *ugrizy* in year 10, respectively. Note that the [1.0x, 30s u] strategy is the same as baseline v3.4.

(CTIO) historical log, and simulated downtime due to weather and maintenance. OpSim outputs are processed by the Metrics Analysis Framework (MAF; R. L. Jones et al. 155) which computes spatially-varying summary statistics (e.g. median seeing) and derived metrics (e.g. coadded 5σ depths) that are used to assess survey strategy with regards to survey efficiency and various science drivers.

We primarily focus on the LSST Wide Fast Deep (WFD) survey, which excludes areas with high galactic extinction ($E(B-V) > 0.2$). We use maps of extragalactic extinction-corrected coadded 5σ point source depths: $\{u_5, g_5, r_5, i_5, z_5, y_5\}$. These maps are in HEALPix [118] format with $N_{\rm side} = 128$, corresponding to a pixel size of 755 arcmin^2.

To quantify the impact of u-band strategy, we compare a series of recent "baseline" simulations,

- baseline v3.4: fiducial simulation of LSST including throughputs for the Ag-Ag-Ag mirror coatings, using band allocations identical to the most recent Al-Ag-Al fiducial simulation.

- baseline v3.5: an update to baseline v3.4 that includes 10% more visits with 38 second exposures in the u band (compared to previous 30 second exposures), as well as the uniform rolling strategy (see Rubin Observatory Survey Cadence Optimization Committee [252]).

- baseline v3.6: an update to baseline v3.5 that includes a more realistic estimate of observatory downtime in year 1 (8 weeks), reducing the number of visits by $\sim 5\%$, as well as an improved model of mechanical inefficiencies in telescope slewing.

- baseline v4.0: an update to baseline v3.6 that includes minor bugfixes to the year 1 downtime that result in slightly more observation time in year 1.

In addition to these baseline strategies, we analyze a suite of variations to baseline v3.4 that vary only the u-band observing allocation, while simultaneously adjusting the $grizy$

allocations to maintain the 10-year duration of LSST (i.e. increasing u-band observing time necessitates decreasing $grizy$ observing time, which are each decreased by the same proportion). These strategies include increasing the number of u-band visits by 10%, 20%, and 50%, while setting the per-visit exposure time to 30, 38, 45, and 60 seconds. Throughout, we refer to these strategies with names such as [1.1x, 38s u], which refers to the strategy that increases u-band visits by 10% and the per-visit exposure time to 38 seconds. Table 4.2 lists each of these simulated variations, their relative change in overall u-band exposure time, and their impact on median depth in each band. Note that for the same amount of u-band observing time, strategies with longer per-visit exposures yield greater depth in the u band compared to more visits with shorter exposures. This is because noise in the u band is dominated by read noise and not sky background due to the fainter sky background in the ultraviolet. Also note that the [1.0x, 30s u] strategy is the same as baseline v3.4. Baselines v3.5 onward have adopted the [1.1x, 38s u] strategy, partially as a result of the findings presented here, but these simulations are not identical to the baseline v3.4 [1.1x, 38s u] simulation due to the other changes described above.

Finally, in addition to the WFD forecasts we make for every survey strategy, for the baseline v4.0 simulation we also estimate LBG number densities in the LSST Deep Drilling Fields (DDFs): COSMOS, the Extended Chandra Deep Field South (ECDFS), ELAIS S1, XMM-LSS, and the Euclid Deep Field South (EDFS). The DDFs are typically 1.3 magnitudes deeper in each band than the WFD survey, however EDFS is shallower than the others as it receives approximately the same number of visits spread over roughly twice the area. COSMOS, furthermore, has an accelerated schedule to build significant depth within the first 3 years to aid cosmology systematics calibration and low-surface-brightness science.

4.3.2 Simulating IGM extinction for $z \leq 3.5$ Galaxies

For our studies of photo-z accuracy we use a simulated catalog based on the Millennium simulation [280], using the GALFORM semi-analytic galaxy formation model [116] and the lightcone construction techniques described by A. I. Merson et al. [203]. This catalog was

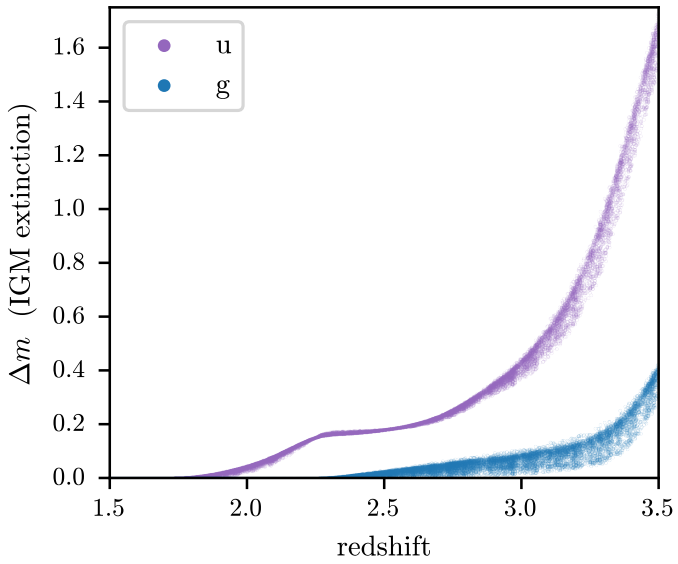

Figure 4.2: IGM magnitude increments in the Rubin u and g bands for $z > 1.5$ galaxies in the simulated catalog. The scatter is due to scatter in UV slope, β_{UV}.

designed to model the optical and near-infrared properties, including emission lines, of $z \leq 3.5$ galaxies detected by LSST. We apply a magnitude cut of $m_i < 25.5$, which is slightly deeper than the DESC gold sample [291] to avoid edge effects in our analysis.

This catalog contains true redshifts and *ugrizy* magnitudes for 240,000 galaxies[4]. The model that generated these true magnitudes includes Lyman-series absorption in the atmospheres of these galaxies, but does not include the effects of extinction in the intergalactic medium (IGM), which also absorbs rest-frame UV flux from these galaxies as photons travel through neutral hydrogen clouds along the line-of-sight. This effect, commonly named the Lyman-alpha forest, is redshift and wavelength dependent, and therefore provides valuable information for photo-z estimation in addition to the information provided by Lyman-series absorption intrinsic to galactic atmospheres. As this information redshifts into the Rubin u

[4]These true magnitudes were computed for the original Al-Ag-Al throughputs. The new throughputs have nearly identical shapes, however, with only the normalization being substantially different. These true magnitudes are, therefore, still valid for the new throughputs.

band at $z > 1.6$ (and the g band at $z > 2.2$), we wish to model this missing IGM absorption to maximize the utility of the u band for photo-z estimation in our simulations.

We use the following model to add IGM extinction at the catalog level. Consider a galaxy with observed-frame SED $f_\lambda(\lambda) \equiv f_\lambda(\lambda, z) \propto f_\lambda(\lambda/(1+z), z=0)$, observed in a bandpass with transmission[5] $R_m(\lambda)$. In the absence of IGM extinction (e.g., in our simulated catalog), we observe the magnitude

$$\begin{aligned} m &= -2.5 \log_{10} f + c_m \\ \text{where} \quad f &= \int f_\lambda(\lambda) R_m(\lambda) \lambda \, d\lambda, \end{aligned} \tag{4.1}$$

and c_m is a band-dependent constant. However, if we include the observed-frame IGM transmission $T(\lambda)$, we observe the magnitude

$$\begin{aligned} m_{\rm wIGM} &= -2.5 \log_{10} f_{\rm wIGM} + c_m \\ \text{where} \quad f_{\rm wIGM} &= \int T(\lambda) f_\lambda(\lambda) R_m(\lambda) \lambda \, d\lambda. \end{aligned} \tag{4.2}$$

Absorption in the IGM, therefore, effectively increments observed magnitudes by the amount

$$\Delta m_{\rm wIGM} = m_{\rm wIGM} - m = -2.5 \log_{10} \frac{f_{\rm wIGM}}{f}. \tag{4.3}$$

We wish to compute these IGM corrections for the galaxies in our catalog.

If we assume the UV SED for each galaxy can be approximated by a power law, $f_\lambda(\lambda) \propto \lambda^{\beta_{\rm UV}}$ where $\beta_{\rm UV}$ is the "UV slope", the flux ratio is then

$$\frac{f_{\rm wIGM}}{f} = \int T(\lambda) \tilde{R}_m(\lambda) \, d\lambda \tag{4.4}$$

where

$$\tilde{R}_m(\lambda) = \frac{\lambda^{\beta_{\rm UV}+1} R_m(\lambda)}{\int \lambda^{\beta_{\rm UV}+1} R_m(\lambda) \, d\lambda}. \tag{4.5}$$

The IGM correction $\Delta m_{\rm wIGM}$, then, depends only on the UV slope $\beta_{\rm UV}$.

[5] This is the dimensionless throughput, giving the probability that a photon with wavelength λ will be detected.

For this simple model, we fit $\beta_{\rm UV}$ from the $u-g$ color of the galaxies. Assuming the u-band flux is the flux of the galaxy at the effective wavelength of the u band,

$$\lambda_u = \frac{\int \lambda\, f_\lambda(\lambda) R_u(\lambda) \lambda\, d\lambda}{\int f_\lambda(\lambda) R_u(\lambda) \lambda\, d\lambda} \tag{4.6}$$

we have

$$u \sim -2.5 \log_{10} \lambda_u^{\beta_{\rm UV}+2} + c_u + c, \tag{4.7}$$

where c is a band-independent constant. Similar equations holds for the g band. We then estimate the UV slope from the $u-g$ color:

$$\beta_{\rm UV} = (u-g) \left(-2.5 \log_{10} \frac{\lambda_u}{\lambda_g} \right)^{-1} - 2. \tag{4.8}$$

Note however, that this definition is circular, as the effective wavelength depends on the UV slope. We, therefore, estimate initial effective wavelengths assuming $\beta_{\rm UV} = -2$, estimate new values for $\beta_{\rm UV}$, and iterate until convergence. For our simulated catalog, final estimates of $\beta_{\rm UV}$ range from -4 for extreme star formers, to 8 for very red galaxies. The bottom of this range matches expectations for extreme star-forming galaxies R. J. Bouwens et al. [30], Y. I. Izotov et al. [149]. We have not found published estimates of UV slopes for red galaxies, as this is a technique typically used for studying star formation. We note, however, that the UV slope provides only a modest modulation to the mean IGM extinction, as seen in Fig. 4.2, so we do not expect this simple model to endanger any conclusions of this chapter.

We use the A. K. Inoue et al. [141] analytic $\tau(\lambda, z)$ model of IGM optical depth to model IGM transmission $T(\lambda, z) = e^{-\tau(\lambda, z)}$. Figure 4.3 plots the three bluest Rubin bandpasses together with the IGM transmission for a few different source redshifts. The galaxies in our simulated catalog are all at $z \leq 3.5$, for which only the u and g band see any significant IGM extinction.

Finally, $\Delta m_{\rm wIGM}$ for galaxies in the simulated catalog are plotted in Figure 4.2. The dispersion in $\Delta m_{\rm wIGM}$ is due to the UV slopes of the galaxies. Whether points above (below) the mean are bluer (redder) than average depends on the source redshift. See Appendix 4.6 for more details.

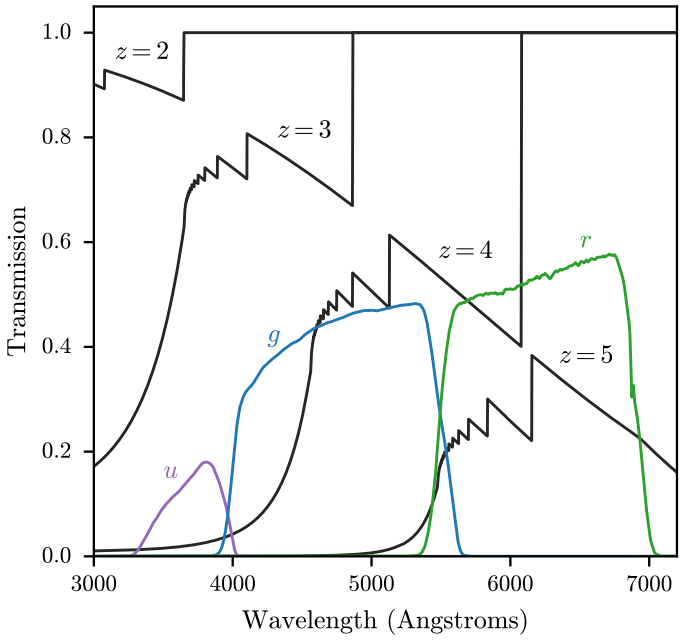

Figure 4.3: Transmission of the IGM in black, plotted for several different source redshifts. The transmission of the LSST photometric bandpasses (with Ag-Ag-Ag coating) are plotted in color to help visualize how much IGM extinction impacts each band at different redshifts.

4.3.3 LBG Flux and Population Model

We model the intrinsic rest-frame spectra of LBGs in absolute magnitudes using a power law,

$$f_\lambda^{\rm intr}(\lambda, M, z) = A \cdot 10^{-0.4M} \left(\frac{\lambda}{1500\,\text{Å}}\right)^{\beta_{\rm UV}(M,z)}, \tag{4.9}$$

where M is the absolute magnitude at 1500 Å[6], and the normalization

$$A = 4.83 \times 10^{-8}\,\text{erg s}^{-1}\,\text{cm}^{-2}\,\text{Å}^{-1}. \tag{4.10}$$

[6]I.e., M is the AB magnitude as calculated with the bandpass $R(\lambda) = \delta(\lambda - 1500\text{Å})$, where δ is the Dirac delta function, assuming the galaxy is at 10 pc (for which redshift $z \sim 0$).

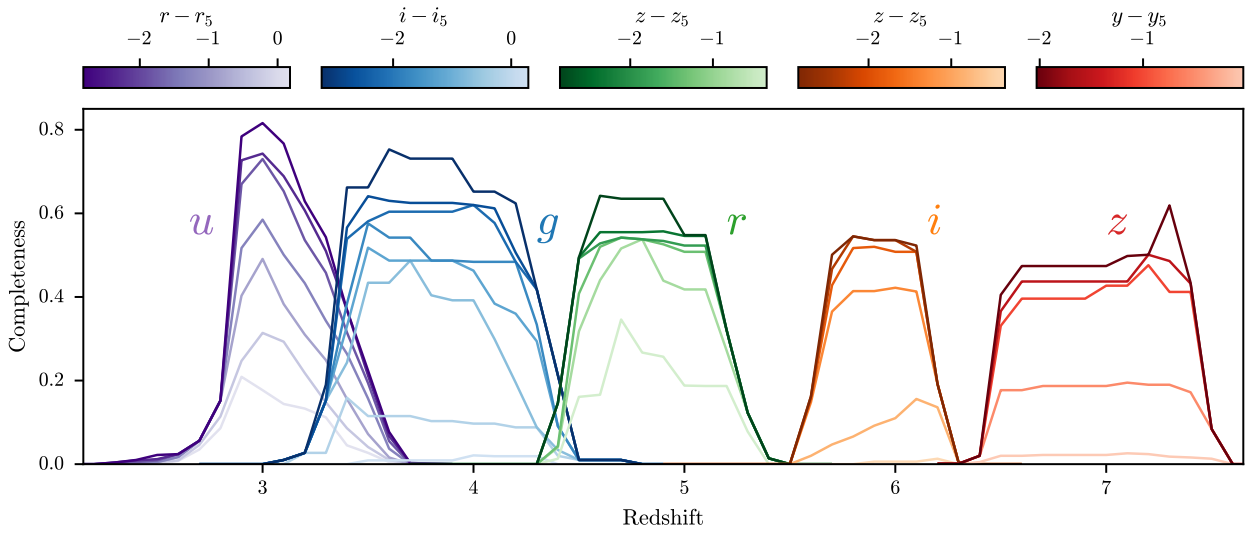

Figure 4.4: LBG completeness curves for *ugriz* dropouts (labeled in the main panel) as a function of redshift, in bins of LBG magnitude relative to the 5σ depth in the detection band (color coded according to the bars above the main panel). u-dropout completeness comes from [194]; *griz* completeness comes from [129, 223].

The UV slope is calculated using the bilinear model

$$\beta_{\rm UV}(M, z) = -0.167(M + 19.5) - 0.063\, z - 1.61, \tag{4.11}$$

which is fit to high-redshift Hubble data from R. J. Bouwens et al. [30] (see Appendix 4.6). The observed spectrum in apparent magnitudes, accounting for redshift and IGM extinction, is then

$$f_\lambda^{\rm obs}(\lambda, M, z) = \frac{1}{1+z}\left(\frac{10\,{\rm pc}}{D_L(z)}\right)^2 f_\lambda^{\rm intr}(\lambda_e, M, z)\, T(\lambda, z) \tag{4.12}$$

where $D_L(z)$ is the luminosity distance to redshift z, and $\lambda_e = \lambda/(1+z)$ is the emitted wavelength. Finally, the observed bandpass magnitude is

$$m(M, z) = -2.5 \log_{10}\left(\frac{\int f_\lambda^{\rm obs}(\lambda, M, z) R_m(\lambda)\lambda\, d\lambda}{\int f_\lambda^{\rm AB}(\lambda) R_m(\lambda)\lambda\, d\lambda}\right), \tag{4.13}$$

where $f_\lambda^{\mathrm{AB}}(\lambda) = 0.109\,(\lambda/\text{Å})^{-2}\,\mathrm{erg\ s^{-1}\,cm^{-2}\,Å^{-1}}$ is the AB reference spectrum (i.e. $f_\nu = 3631\,\mathrm{Jy}$). Note that Equations 4.9-4.13 provide a fully-determined analytic model for m, the apparent magnitude in bandpass R_m, as a function of redshift, z, and the absolute magnitude at rest-frame 1500 Å, M. Alternatively, by numerically inverting these equations, we have a fully-determined model for M as a function of z and m.

True LBG number density as a function of redshift and absolute magnitude at 1500 Å (i.e., the luminosity function) is modeled using a double power law:

$$\phi(M,z) = \phi^* \big[10^{0.4(\alpha+1)(M-M^*)} + 10^{0.4(\beta+1)(M-M^*)}\big]^{-1}, \tag{4.14}$$

where ϕ^* is the characteristic number density, M^* is the characteristic magnitude, β is the bright-end slope, and α is the faint-end slope. For the redshift evolution of each parameter we use the following model:

$$\begin{aligned} \log \phi^*(z) &= -1.45 - 0.31(1+z) \\ M^*(z) &= -21.18 + 0.02(1+z) \\ \alpha(z) &= -1.27 - 0.11(1+z) \\ \beta(z) &= -4.79 + 0.05(1+z). \end{aligned} \tag{4.15}$$

These coefficients, from Table 2 of S. L. Finkelstein & M. B. Bagley [104], are fit to a collection of data from CANDELS [31, 105, 225], the Hubble Frontier Fields [29], HSC SSP [129], CLAUDS [211], SHELA [284], and UltraVISTA/VIDEO [3].

The fraction of true LBGs detected and classified as such by Rubin, as a function of redshift, observed magnitude in the detection band, and 5σ limiting depth in the detection band, is estimated using the LBG completeness models of Y. Harikane et al. [129], M. A. Malkan et al. [194], Y. Ono et al. [223]. In particular, the completeness is quantified in terms of $m - m_5$, the magnitude in the detection band, relative to the 5σ depth in the same band. For *ugriz*-dropout samples, we define the detection band as the Rubin bandpass closest to the redshifted $(1+z)1500$ Å: *rizzy*, respectively. Completeness models, for a discrete set of

$m - m_5$ values in the detection band, are shown in Figure 4.4, from which it is clear that the dropout samples are less complete for galaxies with detection band magnitudes closer to the 5σ limit. For more details, see Appendix 4.6.

Finally, we estimate detected LBG number densities by integrating the luminosity function and completeness model:

$$n = \int_0^\infty dz \, \frac{dV}{dz} \int_{-\infty}^{M_{\rm cut}} dM \, \phi(M, z) \, C(M, z; M_5). \tag{4.16}$$

The absolute-magnitude cut, $M_{\rm cut} = M_{\rm cut}(m_{\rm cut}, z)$, and the absolute-magnitude 5σ depth, $M_5 = M_5(m_5, z)$, are calculated from the detection-band apparent magnitude equivalents, $m_{\rm cut}$ and m_5, by inverting Equation 4.13. Note that m_5 is the apparent magnitude corresponding to a flux with signal-to-noise ratio (SNR) of 5 in the detection band, which is provided by OpSim (see Section 4.3.1), while $m_{\rm cut}$ is the cut on apparent magnitude for LBGs selected for cosmology (see Section 4.4.2). Number densities as a function of detection-band m_5 (where we have set $m_{\rm cut} = m_5$) are shown in Figure 4.5. Median depths for the baseline v4.0 survey strategy are marked, using bars, dots, and stars for WFD years 1 and 10, and COSMOS year 10, respectively.

We do not model contamination from low-redshift interlopers for the metrics presented in this work. Low-redshift interlopers, however, are very important for cosmology and astrophysics applications, and are discussed in Section 4.6.

4.4 Methods

In this section we describe how we use the OpSim simulations of the LSST survey strategy to estimate photo-z's for the simulated Millenium/GALFORM $z \leq 3.5$ catalog, and high-redshift LBG number densities using the analytic population and flux model.

4.4.1 Photo-z Estimation

To estimate photo-z's for the $z \leq 3.5$ simulated catalog we use the Color-Matched Nearest-Neighbors (CMNN) algorithm [119, 120]. CMNN is not chosen because it is the best photo-z

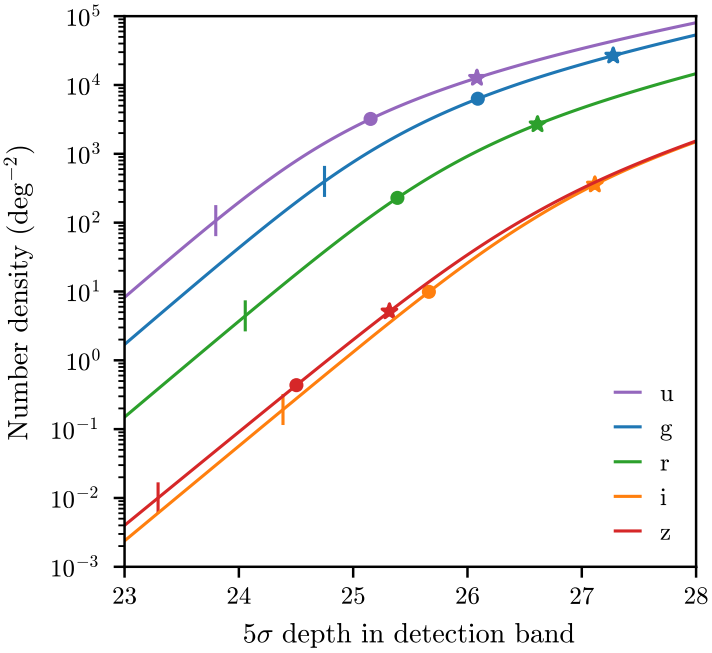

Figure 4.5: Projected number densities for *ugriz* dropouts as a function of magnitude cut in the detection band, assuming $m_{\rm cut} = m_5$. Expected depths and number densities for the baseline v4.0 survey strategy are indicated by bars, dots, and stars for LSST WFD years 1 and 10, and COSMOS year 10, respectively.

estimator, but rather because the accuracy and precision of CMNN estimates are straightforwardly related to the precision of the input photometry. This attribute makes the CMNN photo-z estimator useful for evaluating the relative change in photo-z performance due to varying photometric quality in different survey strategy simulations.

The CMNN estimator is described fully in M. L. Graham et al. [119, 120]. To briefly summarize, it takes as input only two things: the median 5σ depths in each of the *ugrizy* bands for a given survey simulation, and the simulated galaxy catalog of true apparent magnitudes. Given the depths, it calculates observational magnitude uncertainties for every galaxy (using the error model described in J. F. Crenshaw et al. 70, Ž. Ivezić et al. 146),

and then uses the errors to simulate observed apparent magnitudes (i.e., adds randomly generated noise to the true flux). The catalog is then split into a training set of 200,000 galaxies and a test set of 40,000 galaxies, and the training set is used to estimate a photo-z point estimate for each galaxy in the test set.

4.4.2 Forecasting Dropout Number Densities

Constraining large-scale structure with LBGs requires assembling a relatively uniform, high-number density sample across the largest possible area of the sky. These criteria are in tension, as uniformity and number density encourage selection of the deepest areas of the survey, while sky area encourages selecting the widest possible area of the survey (cf. the trade-offs in V. Ruhlmann-Kleider et al. 253 and C. Payerne et al. 226). This trade-off between width and depth will need to be optimized for real analyses that seek to maximize cosmological constraining power. Furthermore, LBG selection criteria, which rely on colors that straddle the observed-frame Lyman-break at $\lambda_{\rm obs} \sim 912\,\text{Å}(1+z)$, will need to be optimized to balance sample size, completeness, and purity according to the needs of cosmology analyses [312]. Here, we make a simple set of choices that enables consistent comparison between different observing strategies:

- we use the deepest 75% of the WFD footprint (corresponding to a sky fraction $f_{\rm sky} = 0.32$);
- in the dropout band, magnitudes below the 3σ limit are replaced by 3σ lower bounds (i.e., magnitudes $m > m_3$ are replaced with m_3, which is understood as a lower bound on the magnitude);
- we require that every LBG used for cosmology have SNR > 5 in the detection band;
- we set the dropout threshold to 1 magnitude; i.e. $m_{\rm dropout} - m_{\rm detection} \geq 1$.

These requirements are used to determine $m_{\rm cut}$ (defined in Section 4.3.3) as a function of the OpSim map for each survey strategy.

For u- and g-dropouts, the dropout band is not sufficiently deep[7] with respect to the detection band (r and i, respectively), so that the dropout band is the limiting factor when determining magnitude cuts. For these samples, therefore, we select the deepest 75% of pixels in the *dropout band* for each OpSim map, and set the LBG magnitude cuts:

$$\begin{aligned} u\text{-drop.:}\ \ r_{\rm cut} &= \min(u_3) - 1 = \min(u_5) - 0.45 \\ g\text{-drop.:}\ \ i_{\rm cut} &= \min(g_3) - 1 = \min(g_5) - 0.45, \end{aligned} \tag{4.17}$$

where we have used

$$m_3 = m_5 + 2.5\log_{10}\frac{5}{3} \approx m_5 + 0.55, \tag{4.18}$$

and the minimum is taken over pixels in the OpSim map.

For riz-dropouts, the dropout band is deeper than the detection band (zzy, respectively). For these samples, therefore, we select the deepest 75% of pixels in the *detection band* for each OpSim map, and set the LBG magnitude cuts:

$$\begin{aligned} r\text{-dropouts:}\quad & z_{\rm cut} = \min(z_5) \\ i\text{-dropouts:}\quad & y_{\rm cut} = \min(y_5) \\ z\text{-dropouts:}\quad & y_{\rm cut} = \min(y_5). \end{aligned} \tag{4.19}$$

[7]Equation 4.17 below makes clear the criterion for being "sufficiently deep" is that the 5σ depth in the dropout band is more than 0.45 magnitudes deeper than the 5σ depth in the detection band. See Table 4.2 for the relative depth of each band.

For example, the cuts for the baseline v3.6 survey strategy in year 1 (10) are

$$\begin{aligned} u\text{-dropouts:} \quad & r_{\rm cut} = 23.48\ (24.72) \\ g\text{-dropouts:} \quad & i_{\rm cut} = 24.64\ (26.00) \\ r\text{-dropouts:} \quad & z_{\rm cut} = 24.18\ (25.57) \\ i\text{-dropouts:} \quad & y_{\rm cut} = 24.18\ (25.57) \\ z\text{-dropouts:} \quad & y_{\rm cut} = 23.33\ (24.63), \end{aligned}$$

where, again, the minimum is taken over pixels in the OpSim map.

Using these values for $m_{\rm cut}$, we evaluate Equation 4.16 across the OpSim map (keeping only the deepest 75% of pixels) and take the mean. This mean density is the metric used to compare LBG detection for different survey strategies. Note that while $m_{\rm cut}$ is set for the entire sample, m_5 varies spatially across the survey footprint due to spatial variation in observing quality. Even within this deepest 75%, non-uniformity of LBG number densities due to spatial modulation of m_5 will be important for studies of large-scale structure, as this modulation in number density can be mistaken as large scale structure [308]. We will investigate the impact of this modulation on forecasts for LSST cosmology in future work. We also do not model contamination by low-redshift interlopers, the impact of which are discussed in Section 4.6.

4.5 Results

4.5.1 Photo-z Estimation

We use the CMNN algorithm to estimate photo-z's for every galaxy in the simulated $z \leq 3.5$ catalog, using OpSim maps of *ugrizy* 5σ depths for LSST year 10. We quantify photo-z accuracy via the quantity $\Delta z = (z_{\rm phot} - z_{\rm true})/(1 + z_{\rm true})$, the numerator of which quantifies the photo-z error, while the denominator compensates for the larger uncertainty at high redshifts. We then bin galaxies by true redshift and calculate the following quantities for each bin:

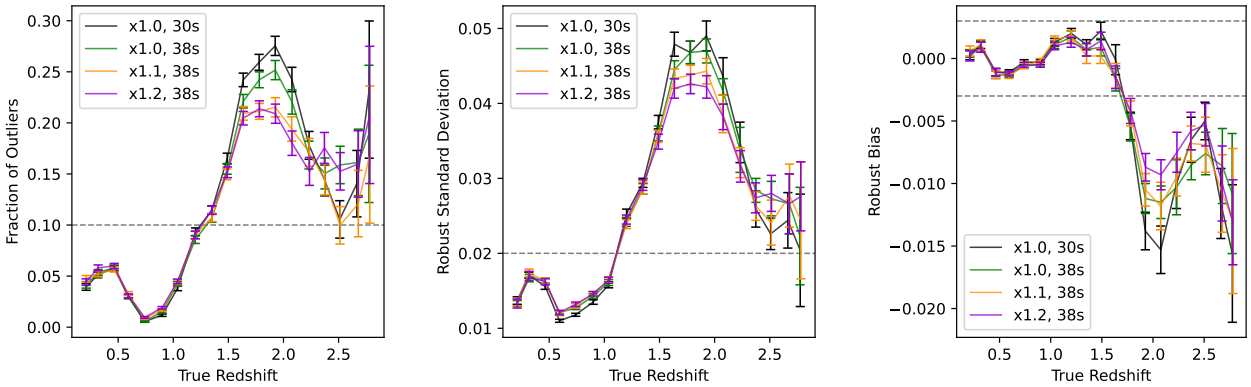

Figure 4.6: Photo-z metrics for simulations of several different u-band survey strategies. For clarity, we plot only a small subset of the u-band strategies considered. Each panel displays the corresponding requirement for LSST science as a horizontal gray line [144]. These requirements provide a sense of scale for each metric, but whether or not this photo-z estimator achieves each limit in different redshift ranges is not predictive of the photo-z performance of DESC cosmology, due to the considerations discussed in Section 4.4.1.

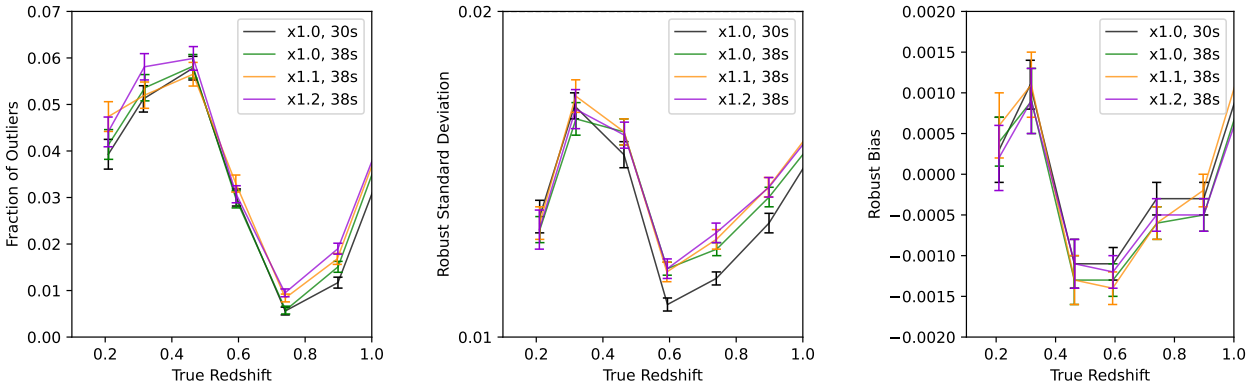

Figure 4.7: Same as Fig. 4.6, zoomed in on the $z < 1$ region for each metric.

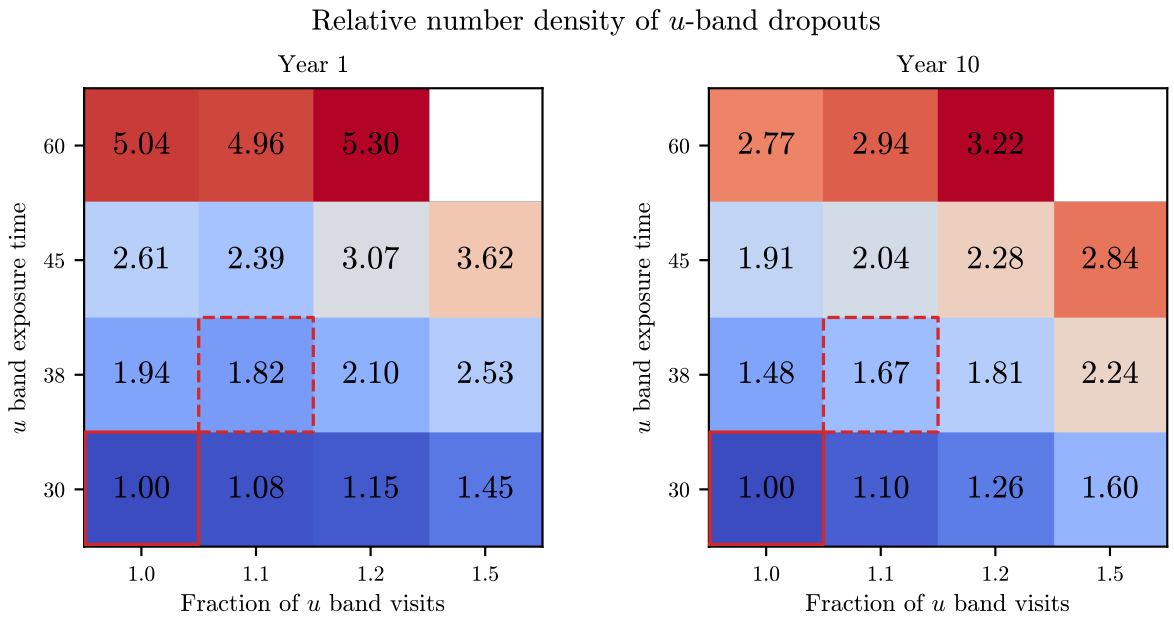

Figure 4.8: Relative number density of $z \sim 3$ u-band dropouts detected across the deepest 75% of the survey footprint for different simulations of u-band survey strategy. Left panel corresponds to LSST year 1, right panel to year 10. The block corresponding to baseline v3.4, [1.0x, 30s u], is bordered by a solid red box, while the SCOC recommendation of [1.1x, 38s u] is bordered by a dashed red box. The absolute number density for baseline v3.4 is $69\,\mathrm{deg}^{-2}$ in year 1 and $2113\,\mathrm{deg}^{-2}$ in year 10.

- the robust standard deviation, $\sigma_{\Delta z}$, which we define as the width of the interquartile range (IQR) of Δz, divided by 1.349 to convert to the equivalent of a Gaussian standard deviation;
- the photo-z bias, which we define as the mean value of Δz for galaxies within the IQR;
- the outlier fraction, which we define as the fraction of galaxies for which $|\Delta z| > 3\sigma_{\Delta z}$.

When calculating the first two quantities, we exclude the galaxies that are flagged as outliers, so these two quantities characterize the core of the distribution, while the outlier fraction

characterizes the tails.

These quantities, as a function of redshift, are plotted in Fig. 4.6, with the colors corresponding to different u-band strategies (recall Section 4.3.1 and Table 4.2). All three metrics significantly degrade above $z \sim 1.2$, at which point the Balmer break redshifts out of the Rubin bandpasses and there is relatively little information present in galaxy spectra for broadband photo-z estimation, resulting in significantly degraded photo-z accuracy [156, 189]. Around $z \sim 1.6$, however, Lyman-series transitions (i.e., the Lyman-alpha forest) begin to redshift into the Rubin u band, stealing progressively more of the u-band flux. This provides a distinctive signal for photo-z estimation, resulting in a reversal of the trend, with photo-z accuracy improving until about $z \sim 2.5$. Our simulations contain very few galaxies beyond this redshift, resulting in increased photo-z errors and uncertainties in the quantities plotted in Fig. 4.6.

For all three quantities, increasing the total u-band exposure time improves photo-z results in the range $1.5 < z < 2.5$: the outlier fraction improves by up to 30%; the standard deviation improves by up to 20%; the bias improves by up to 40%. These improvements are due to the greater u-band depth increasing the SNR of the u-band flux decrement that results from the redshifting of Lyman-series absorption into the Rubin u band.

The improvements in photo-z estimation at $1.5 < z < 2.5$, however, come with a loss of performance at $z < 1$ due to the decreased depth in the *grizy* bands (because more time spent observing in u must result in less time observing in *grizy*, due to the fixed 10-year duration of LSST). The three panels of Fig. 4.7 show the same quantities as Fig. 4.6, zoomed in on the region $z < 1$. These losses, while smaller in magnitude than the gains at $1.5 < z < 2.5$, are, however, statistically significant, especially for the outlier fraction and scatter, which both degrade by up to 3σ. It must also be kept in mind that the $z < 1$ galaxy sample contains far more galaxies, both in our simulations and for the future LSST survey. Thus these small performances losses at $z < 1$ may be judged to outweigh the gains at $1.5 < z < 2.5$, depending on the science case in consideration.

We note, however, that our simulated photometry, while noisy, is free of biases, aperture

corrections, and other systematic errors. These effects, present in real measured photometry, are likely to erase the small changes in photo-z performance at $z < 1$ that result from the small changes in *grizy* depth (cf. Table 4.2). We judge, therefore, that the impressive gains at $z > 1.5$ provide strong motivation for increasing u-band observing time. Indeed, baselines v3.5 onward have adopted the [1.1x, 38s u] strategy, partially as a result of the findings presented here.

4.5.2 LBG Detection

We forecast LBG detection for each survey simulation using the strategy described in Section 4.4.2, considering first only u-band dropouts, which provide LBG samples at $z \sim 3$. We discuss the higher-redshift *griz* dropouts at the end of this section.

For the deepest 75% of the survey footprint in baseline v3.4, we forecast a u-band dropout number density of $69\,\mathrm{deg}^{-2}$ in year 1 and $2113\,\mathrm{deg}^{-2}$ in year 10. Relative LBG number densities for the range of u-band strategy simulations are displayed in Fig. 4.8. Increasing u-band depth by increasing the number of u-band visits and/or the u-band per-visit exposure time results in a greater number density of detected LBGs due to the correspondingly deeper cut that is allowed in the detection band (the r band; cf. Equation 4.17).

Note that for a fixed increase in total u-band observing time, increasing the per-visit exposure time has a greater impact than increasing the number of u-band visits. For example, the [1.5x, 30s u] and [1.0x, 45s u] both correspond to increasing the total u-band observing time by 50%, but the latter strategy with longer per-visit exposures results in a much greater increase in LBG number density. This is because the Rubin u band is read-noise limited due to the lower sky background at these wavelengths. Indeed, the impact of a 10% increase in the number of u-band visits is so marginal that the potential for extra accumulated u-band depth is not sufficient by the end of year 1 to overcome natural depth variations between different simulation realizations. Thus, the "1.1" column in the left panel of Fig. 4.8 reports lower LBG number densities than the "1.0" column. This discrepancy also provides an estimate for the precision of our estimated LBG number densities, when considering uncertainties due

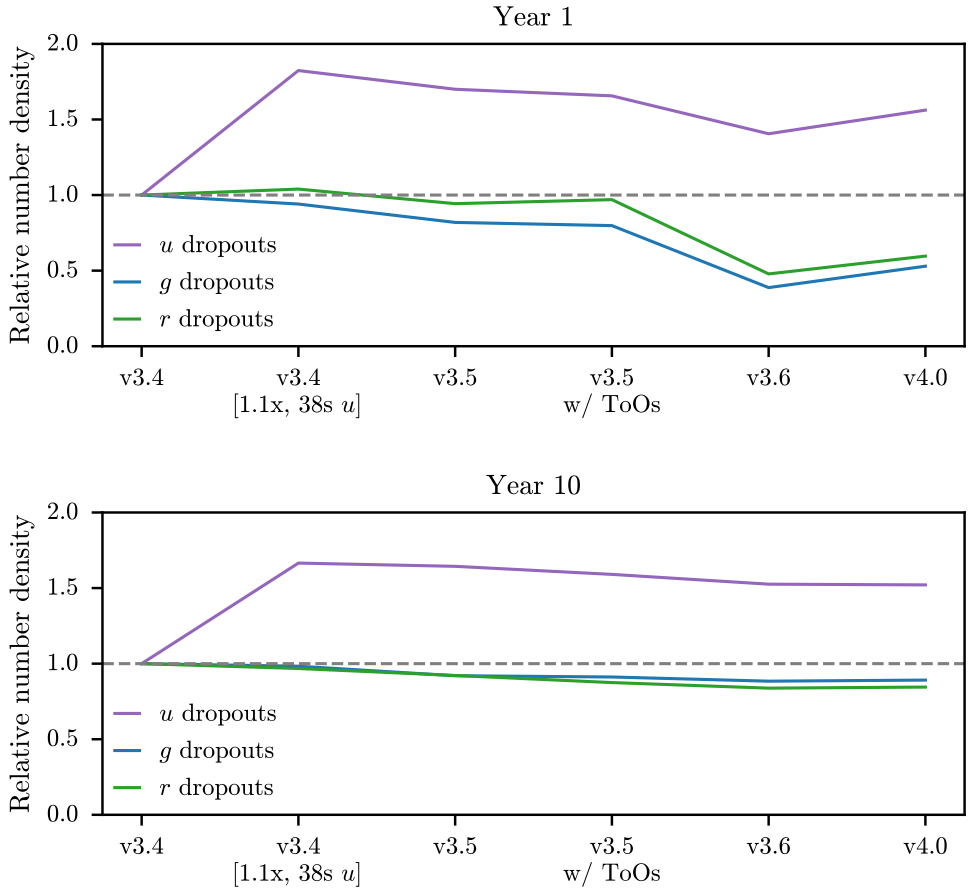

Figure 4.9: Evolution of u-, g-, and r-band dropouts for a series of subsequent survey simulations. v3.4 is the baseline simulation after the adoption of the new Ag-Ag-Ag mirror coatings. v3.4 [1.1x, 38s u] is the same as v3.4, except with 10% more visits in the u band, with a per-visit exposure time of 38 seconds (compared to 30 seconds in baseline v3.4). v3.5 is the new baseline after the adoption of the [1.1x, 38s u] strategy, and includes a few other changes. v3.5 w/ ToOs is the same as baseline v3.5, except with some survey time dedicated to Target of Opportunities. Baselines v3.6 and v4.0 include more realistic simulations of year 1 observatory downtime.

Field	n_u	n_g	n_r	n_i	n_z
WFD	110 (3200)	400 (6300)	4.4 (230)	0.19 (9.9)	0.01 (0.44)
COSMOS	9400 (13000)	13000 (27000)	1100 (2700)	120 (360)	0.74 (5.1)
ECDFS	5400 (13000)	6900 (23000)	260 (2300)	23 (290)	0.25 (4.3)
ELAIS S1	3000 (13000)	4700 (22000)	340 (1700)	29 (210)	0.17 (2.0)
XMM-LSS	4400 (11000)	4400 (18000)	220 (1600)	19 (190)	0.14 (2.3)
EDFS	2000 (15000)	5400 (25000)	280 (2400)	20 (310)	0.23 (3.8)

Table 4.3: Forecast LBG number densities for year 1 (10) of LSST assuming survey strategy baseline v4.0. All number densities are in units of deg^{-2}, rounded to 2 significant digits.

to natural variations between different simulation realizations.

Another important feature to recognize is that the proposed increases to u-band depth have a larger impact in year 1 than in year 10 when comparing relative number of LBGs detected. Compare, for example, the relative LBG number density increase for the most aggressive u-band strategy, which increases the number of u-band visits by 20% and the u-band exposure time to 60 seconds. In year 10, this strategy results in 3.2× more LBGs detected, a large increase that is, however, much smaller than the corresponding 5.3× increase in year 1. This is because year 1 depths are on a steeper part of the LBG luminosity function compared to the deeper year 10 depths (cf. Fig. 4.5), so the increase in LBG number density per unit depth is greater in year 1 than in year 10.

In October 2024 the Rubin SCOC recommended that LSST adopt the [1.1x, 38s u] strategy for the u band [252]. Note with this strategy and the Ag-Ag-Ag mirror coatings, all bands are deeper than under the previous Al-Ag-Al baseline. This strategy, marked by the dashed-red boxes in Fig. 4.8, was incorporated into the new baseline simulation, v3.5. Baseline simulations v3.5, v3.6, and v4.0 include other changes, unrelated to u-band strategy, that also impact forecast LBG number densities, including dedicating survey time to targets

of opportunity[8] (ToOs), as well as increasing the amount of year 1 observatory down-time to more realistic levels. The impacts of these changes to survey strategy on the detection of LBGs are shown in Fig. 4.9.

In year 1, the number of detected u-band dropouts is seen to increase by a factor of 1.8 under strategy [1.1x, 38s u], an increase that is mostly sustained across baseline v3.5 and the subsequent inclusion of ToOs. There is a significant decrease in u-band dropout detection in baseline v3.6 due to the more realistic amount of observatory downtime simulated in year 1. This decrease is somewhat mitigated by the improvements to simulating year 1 downtime implemented in baseline v4.0 (see Section 4.3.1). A similar pattern is visible for year 10 detections, however the decrease in LBG number densities due to the increased year 1 downtime is far smaller, due to this comprising a much smaller fraction of survey time by year 10.

In both panels we also plot the forecast number density of g- and r-band dropouts, at redshifts $z \sim 4$ and $z \sim 5$, respectively. It is seen that the chosen increase to u-band survey time results in only very modest reductions in the number densities of these higher-redshift LBGs. While not shown in the plots, the same is true of z- and y-band dropouts, at redshifts $z \sim 6$ and $z \sim 7$, respectively. Number densities for all dropout samples for the baseline v4.0 survey strategy are shown in Fig. 4.5.

Finally, we list projected LBG number densities for simulation baseline v4.0 in Table 4.3. Densities are listed for years 1 and 10, including the WFD survey and all LSST deep fields. Note that densities are much greater in COSMOS in year 1, compared to the other deep fields, reflecting the early emphasis on building depth in COSMOS for photo-z calibration and low-surface-brightness science.

[8]ToOs are transient events, such as gravitational wave detections [66], that require immediate follow-up observations.

4.6 Conclusions

This chapter used OpSim simulations of the LSST survey, together with simple models of IGM absorption and LBG dropout detection, to evaluate the impact of LSST u-band observing strategy on photo-z estimation and detection of LBGs. We find that:

- Adjusting LSST strategy to increase u-band depth has a small, but statistically-significant negative impact on photo-z estimation for galaxies at $z < 1.5$. We expect these small changes in performance, however, would be erased by a more realistic treatment of photometry that includes systematic errors, such as aperture corrections.

- For galaxies at $z > 1.5$, increasing the u-band depth yields a significant improvement in photo-z estimation.

- Increasing u-band depth has the potential to dramatically increase the number of $z \sim 3$ u-band dropouts detected by LSST.

These metrics were presented to the Rubin SCOC in June 2024, motivating the SCOC to recommend increasing the number of u-band visits by 10% and the per-visit exposure time to 38 seconds [252], which has been adopted as the baseline for survey strategy simulations going forward. With this strategy and the Ag-Ag-Ag mirror coatings, the projected depths in all six bands are deeper than the previous Al-Ag-Al baseline.

With the adoption of this u-band strategy, our metrics indicate that

- The outlier fraction, standard deviation, and bias of $z < 1.5$ galaxies will *increase* by 20%, 5%, and 50%, respectively. These values are, however, small in absolute magnitude, and all three metrics remain well below the LSST requirements.

- The outlier fraction, standard deviation, and bias of $z > 1.5$ galaxies will *decrease* by up to 28%, 10%, and 25%, respectively. As all three metrics are significantly larger at high-redshifts, these also represent large improvements in absolute value.

- The number density of u-band dropouts will increase by 82% in year 1 and 67% in year 10. Forecast number densities for $ugriz$-dropouts in the LSST WFD and DDFs are listed in Table 4.3.

This represents significant gains for science with high-redshift galaxies. For example, the u-band dropout sample forecast to have number densities of 110 deg^{-2} (3200 deg^{-2}) in year 1 (10) will enable measurement of the cross-correlation with Simons Observatory CMB lensing at an SNR of 90 (160) [4, 312]. Such powerful constraints on high-redshift large-scale structure will enable 1% constraints on the high-redshift evolution of σ_8 [312], competitive constraints on the amplitude of local-type primordial non-Gaussianity [48, 269], optical-depth-independent constraints on the sum of neutrino masses [320], and high-precision constraints on the masses of $z > 1$ clusters [301].

Given the high SNR forecast for our observables, the precision of cosmological constraints will ultimately be limited by our control of systematic errors. One of the most important sources for systematic error to consider is photo-z contamination from low-redshift interlopers. Galaxies with strong Balmer/4000 Å breaks, dusty galaxies, emission line galaxies (ELGs), and low-temperature dwarf stars all contaminate LBG dropout samples because they have colors that mimic the Lyman break [223, 237, 281, 307]. Indeed, V. Ruhlmann-Kleider et al. [253] and C. Payerne et al. [226] found the purity of u-band dropout samples depends strongly on u-band depth, with the number of low-redshift interlopers dropping by 38% as u-band depth increases from 24.5 to 25.5. Furthermore, deep u-band imaging is valuable for suppressing interlopers in high-redshift dropout populations by rejecting sources with non-negligible flux in wavelengths bluer than the supposed dropout band [307]. Thus, while this chapter considers only detected number densities of true LBGs, greater u-band depth will also increase purity for LBG dropout samples at all redshifts. The deep photometry and data at other wavelengths available in the LSST DDFs will be invaluable for characterizing interloper populations present in LSST LBG dropout samples. Careful study and calibration of these populations will be necessary to enable precision cosmology with

LBGs detected by LSST.

Finally, we note that increasing the u-band depth of LSST imaging enhances synergies with the proposed DESI-II survey [265], the extension of Dark Energy Spectroscopic Instrument (DESI) [83] which aims to map the 3D matter distribution in the $2 < z < 4$ universe using spectroscopic samples of LBGs and Lyman-alpha emitters (LAEs), enabling tests of cosmological models in the matter-dominated era. DESI-II plans to use LSST year 2 catalogs to select candidate LBGs and LAEs for spectroscopic follow-up. As previously discussed, increasing LSST u-band depth will increase the number density and purity of these candidate samples, increasing the efficiency of the DESI-II survey.

The code to produce the plots in this chapter are available on GitHub[9]. The code for the LBG models described in Section 4.3.3 are published as the python package `lbg_tools`[10], which is available on PyPI[11].

Appendix 4.a: Comments on β_{UV} dependence of IGM increments Δm_{wIGM}

Note the shapes of IMG transmission curves in Figure 4.3 provide some insight into the effect of the UV slope. When the Lyman-alpha forest starts to redshift into a band, bluer spectra will have a larger IGM correction, $\Delta m_{\rm IGM}$, as they have greater flux in the wavelength range impacted by extinction. However, as IGM extinction redshifts farther into the band, redder spectra will have a larger correction, as the IGM has a greater optical depth at high redshift, corresponding to longer wavelengths. You can see this from the deeper troughs on the right side of the IGM extinction curves in Figure 4.3. However, once the Lyman limit redshifts into the band, bluer spectra will once again have larger corrections as the IGM is far more opaque at wavelengths below the Lyman limit.

Figure 4.10 plots $\Delta m_{\rm IGM}$ for the u- and g-bands as a function of both redshift and UV slope. It is difficult to see the structure described above, however, indicating that these

[9] https://github.com/jfcrenshaw/u-band-strat

[10] https://github.com/jfcrenshaw/lbg_tools

[11] https://pypi.org/project/lbg-tools/

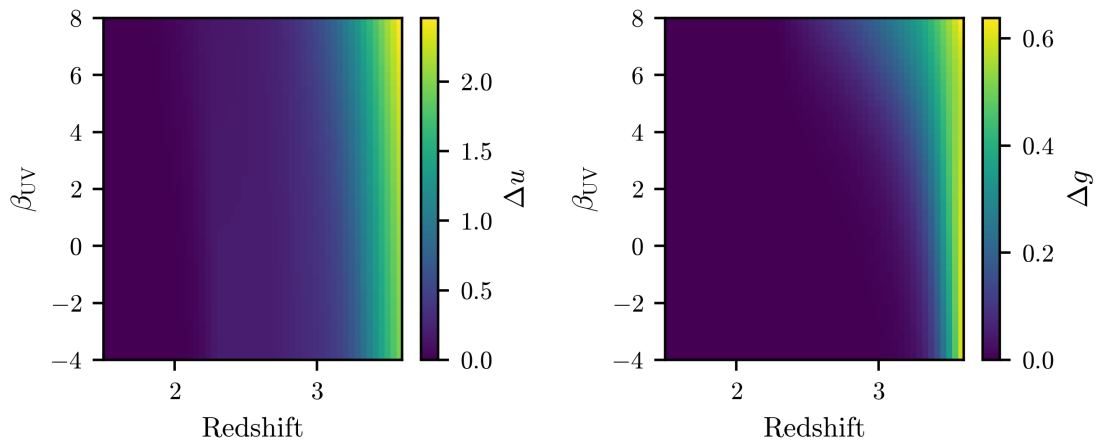

Figure 4.10: IGM magnitude increments for the Rubin u and g bands, as a function of source redshift and UV slope.

effects are relatively small.

Appendix 4.b: Fitting the $\beta_{\rm UV}$ model

To model the UV spectra of LBGs we use HST data from R. J. Bouwens et al. [30]. Specifically, we use the linear fit parameters listed in Table 3 for six tomographic redshift bins between $2.5 \lesssim z \lesssim 8.0$. These parameters, as a function of mean redshift, are plotted in Fig. 4.11.

The parameter $\beta_{\rm UV}$ at $M_{\rm UV} = -19.5$ show a clear linear trend with redshift, so we fit a linear model, yielding the relation

$$\beta_{\rm UV}|_{M_{\rm UV}=-19.5} = -0.063z - 1.61. \tag{4.20}$$

The parameter $\mathrm{d}\beta_{\rm UV}/\mathrm{d}M_{\rm UV}$ does not show a clear trend, so we simply take the average:

$$\left\langle \frac{\mathrm{d}\beta_{\rm UV}}{\mathrm{d}M_{\rm UV}} \right\rangle = -0.167. \tag{4.21}$$

Together, these two relations yield the bilinear model

$$\beta_{\rm UV}(M, z) = -0.167(M + 19.5) - 0.063\, z - 1.61, \tag{4.22}$$

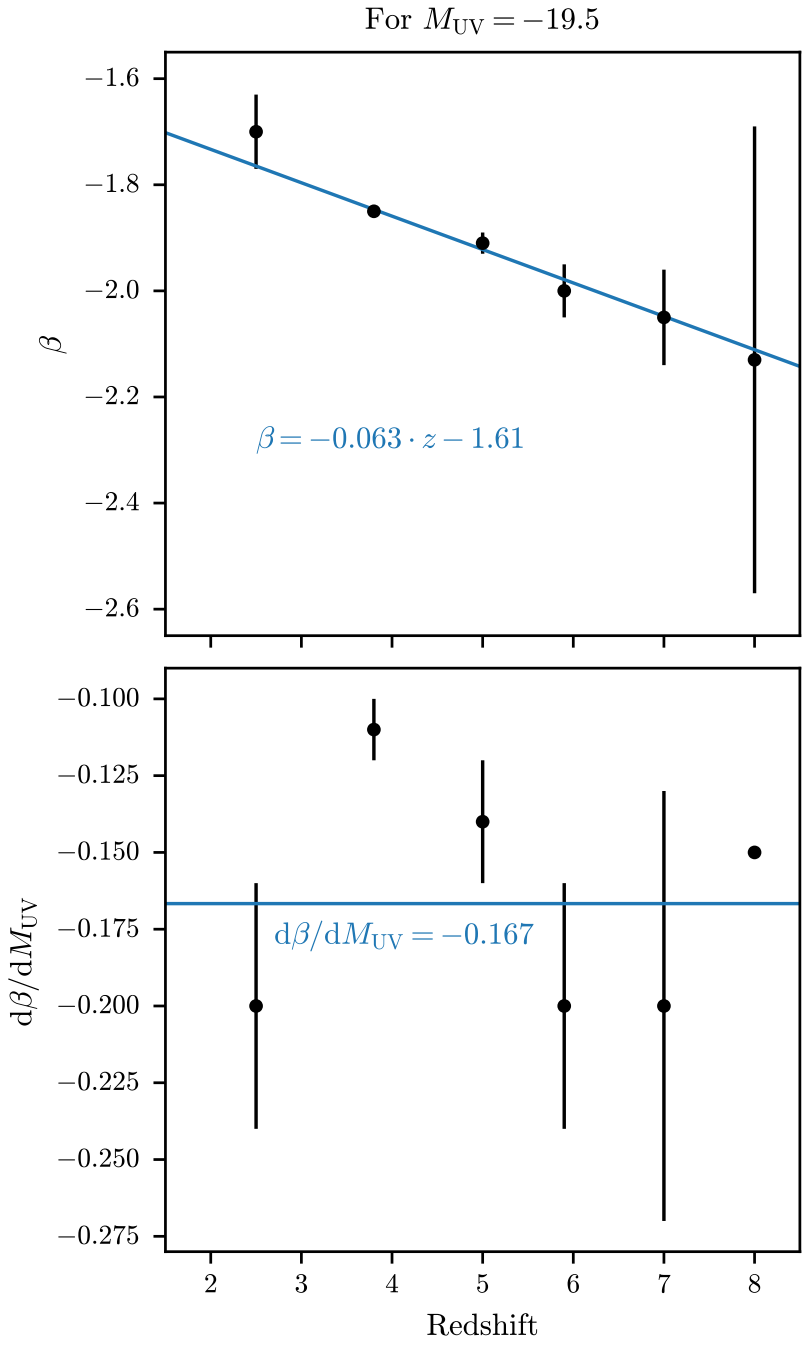

Figure 4.11: Model for evolution of β_{UV}, fit to data from Table 3 of [30].

which is also printed as Eq. 4.11.

Appendix 4.c: Details of the LBG Completeness Model

The LBG completeness models of Y. Harikane et al. [129], M. A. Malkan et al. [194], Y. Ono et al. [223] were calibrated using synthetic source injection (SSI). That is, synthetic galaxy images with a variety of spectral types, redshifts, and intrinsic magnitudes were injected into real images, and these images were processed using the usual science pipelines used for these studies. The fraction of true LBGs injected into the images were then compared with the number of LBGs that were detected in the images and then passed the corresponding color cuts. The completeness was then calculated in bins of redshift and apparent magnitude.

Estimates for u-dropout completeness come from M. A. Malkan et al. [194]. This study used photometry from the Subaru Deep Field, and detected LBGs using z-band imaging with 5σ depth 26.07. While we use these same completeness curves, we assume u-dropout detection occurs in the r-band, which for LSST has a 5σ depth ~ 25.7 in year 1 and ~ 26.8 in year 10, which is well-matched to the z-band depth from the Subaru Deep Field. Estimates for $griz$-dropout completeness come from Y. Harikane et al. [129], Y. Ono et al. [223], which use photometry from the HSC SSP Wide, Deep, and Ultradeep fields. These fields span a range of 5σ depths that bracket the expected LSST year 1 and 10 depths.

In every case, we use the strategy of Y. Harikane et al. [129] to "rescale" completeness estimates to imaging of different depths. That is, we take the input grid of redshift, apparent magnitude, and completeness values, $\{z,\ m,\ C\}$, and subtract the 5σ depth from the apparent magnitudes:

$$\{z,\ m,\ C\} \rightarrow \{z,\ m - m_5,\ C\}. \tag{4.23}$$

Completeness, therefore, is modeled as a function of redshift and *magnitude relative to the 5σ depth in the detection band*. We then linearly interpolate (and extrapolate) using this grid, with completeness values clipped to the range $[0, 1]$ to keep completeness for very bright (faint) galaxies ≤ 1 (≥ 0). We smooth the input completeness curves by requiring that

interpolated/extrapolated completeness values decrease monotonically with magnitude and is unimodal with respect to redshift. This improves the behavior of extrapolation beyond the input calibration grid. The "smoothed" curves are displayed in Fig 4.4.

Note we do not expect these completeness models to be correct *in detail* for LSST, however we do expect they provide reasonable estimates for the order of magnitude of LBGs that LSST will detect, reasonable redshift ranges for these detections, and a reasonable scaling with LBG magnitude relative to the 5σ depth of the imaging. Quantifying completeness in detail for LSST will be vital for cosmology with LBGs detected by Rubin, however we do not think these details impact the conclusions of this chapter.

Finally, we note that Equation 4.16 uses absolute magnitudes at 1500 Å, M, rather than apparent magnitudes, m. Equations 4.9-4.13, however, provide a fully-determined analytic model for $m = m(M, z)$. These equations can be numerically inverted to provide a fully-determined model for $M = M(m, z)$.

Chapter 5

FORECASTING HIGH-REDSHIFT COSMOLOGY USING LYMAN-BREAK GALAXIES DETECTED BY LSST

In this chapter we forecast cosmology constraints using Lyman-break galaxies detected by LSST. These forecasts set expectations for what we can expect from LBG cosmology, what observables are worth pursuing in year 1 vs year 10 of the survey, and how different systematic errors impact our constraints.

We will consider two so-called 3×2pt analyses: a low- and high-redshift variant. The low-redshift analysis is the standard for LSST cosmology, and consists of correlations between galaxy positions and the weak gravitational lensing of galaxy shapes. The high-redshift analysis provides analogous constraints at high-redshift, this time correlating LBG positions with weak gravitational lensing of the CMB. These 3×2pt analyses provide significant constraining power for cosmology, while simultaneously constraining systematic error models. We will find that the LBG constraints, when considered in isolation, are weaker than the low-redshift constraints, but that they are highly complementary, so that their combination yields significant improvements for cosmology.

In Section 5.1 we describe the Fisher matrix formalism used to generate these forecasts. In Section 5.2 we detail the large scale structure observables that allow us to constrain cosmological parameters, and in Section 5.3 we describe the systematic errors we model for LBGs. In Section 5.4 we calculate the covariance matrix associated with our observables, and in Section 5.5 we optimize the magnitude cuts used to assemble our LBG samples. In Section 5.6 we forecast constraints on cosmological parameters using the ingredients from all of the previous sections, before concluding in Section 5.7.

5.1 Fisher Forecasting

To forecast the constraining power of our high-redshift observables we perform Fisher forecasts [58]. That is, we calculate the Fisher Information matrix, which is the expected curvature of the log-likelihood around it's maximum:

$$F_{ij} = \left\langle -\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j} \right\rangle, \tag{5.1}$$

where L is the likelihood of the data given model parameters θ_i. The covariance of model parameters θ_i, θ_j is then

$$\mathrm{Cov}_{ij} = (F^{-1})_{ij}. \tag{5.2}$$

Assuming the likelihood is approximately Gaussian near its maximum, the Fisher matrix can be calculated

$$F_{ij} = \sum_{\ell,\ell'} \frac{\partial C_\ell}{\partial\theta_i} (\mathrm{Cov}^{-1})_{\ell\ell'} \frac{\partial C_{\ell'}}{\partial\theta_j}, \tag{5.3}$$

where $\mathrm{Cov}_{\ell\ell'}$ is the data covariance and C_ℓ is the theory prediction for the observable signal.

Notice that Fisher forecasts depend only on a method to estimate the signal C_ℓ and a forecast for the data covariance $\mathrm{Cov}_{\ell\ell'}$. Furthermore, due to the Cramér-Rao lower bound, Fisher forecasts predict the minimum possible uncertainty on unbiased model parameters θ_i as constrained using the provided data and model.

The following few sections detail observables associated with large scale structure, as well as a calculation for their covariance. These ingredients enable a forecast of the corresponding constraints on cosmological parameters.

5.2 Large Scale Structure Observables

In this section we describe the observables utilized in our forecast, which consist of two independent[1] 3×2pt analyses. A 3×2pt analysis combines 3 different two-point functions

[1] Technically these probes are not totally independent due to weak lensing magnification bias. See Section 5.4.

in a complementary manner to break parameter degeneracies, resulting in an analysis that is more robust to systematic errors than an analysis using only a single two-point function.

The first of these is the traditional 3×2pt analysis planned by the DESC, which consists of galaxy clustering, galaxy-galaxy lensing, and cosmic shear of $z \lesssim 2$ galaxies. We refer to this analysis as the low-redshift analysis. We complement this analysis by introducing a high-redshift 3×2pt analysis using Lyman-break galaxies in three tomographic redshift bins, cross-correlated with CMB lensing. This analysis is very similar to the low-redshift analysis, except it replaces tomographic galaxy shears with the shear of the CMB, which lives on a single, well-known, high-redshift source plane.

In the following subsections we introduce the formalism behind these analyses, including a model for intrinsic alignments, which is a very important source of systematic errors in cosmic shear analyses.

5.2.1 Galaxy clustering

Both of our 3×2pt analyses utilize galaxy clustering, i.e. correlations in galaxy positions, to constrain large-scale structure. In the absence of systematic errors, the observed number density of galaxies is the sum

$$\hat{n}_g(\boldsymbol{\theta}) = n_g(\boldsymbol{\theta}) + n_{\rm rnd}(\boldsymbol{\theta}), \tag{5.4}$$

where $n_g(\boldsymbol{\theta})$ is the true number density of galaxies and $n_{\rm rnd}(\boldsymbol{\theta})$ is the Poisson shot noise. We are interested in the statistical properties of this field. It is convenient to work in terms of the over-density, $\delta_g((\theta)) \equiv (n_g(\boldsymbol{\theta}) - \bar{n}_g)/\bar{n}_g$, where $\bar{n}_g$ is the average galaxy number density. We then have

$$\hat{\delta}_g(\boldsymbol{\theta}) = \delta_g(\boldsymbol{\theta}) + \frac{n_{\rm rnd}(\boldsymbol{\theta})}{\bar{n}_g}. \tag{5.5}$$

The two-point correlation function between two galaxy samples i and j is then

$$\langle \hat{\delta}_g^i(\boldsymbol{\theta}) \hat{\delta}_g^j(\boldsymbol{\theta}) \rangle = \langle \delta_g^i(\boldsymbol{\theta}) \delta_g^j(\boldsymbol{\theta}) \rangle + \bar{n}_g^{-2} \langle n_{\rm rnd}^i(\boldsymbol{\theta}) n_{\rm rnd}^j(\boldsymbol{\theta}) \rangle. \tag{5.6}$$

The Fourier transform of this correlation function is

$$\langle \hat{\delta}_g^i(\ell)\hat{\delta}_g^j(\ell')\rangle = (2\pi)^2\delta^{(2)}(\ell+\ell')\hat{C}_{gg}^{ij}(\ell) \tag{5.7}$$

where[2]

$$\hat{C}_{gg}^{ij}(\ell) = C_{gg}^{ij}(\ell) + \bar{n}_g^{-1}\,\delta_{ij}, \tag{5.8}$$

and C_{gg}^{ℓ} is the 2D power spectrum between bins i and j.

The 2D galaxy-galaxy power spectrum can be calculated as the projection of the 3D matter power spectrum, $P_{\delta\delta}(k,z)$, using the Limber approximation [186]:

$$C_{gg}^{ij}(\ell) = \int \mathrm{d}\chi\, \frac{q_g^i(\chi)\, q_g^j(\chi)}{\chi^2} P_{\delta\delta}\left(k = \frac{\ell+\frac{1}{2}}{\chi}, z\right), \tag{5.9}$$

where χ is the comoving distance, $z \equiv z(\chi)$ is the cosmological redshift at distance χ, and the transfer function

$$q_g^i(\chi) = n^i(z)\, b^i(z) \frac{dz}{d\chi}. \tag{5.10}$$

Here $n^i(z)$ and $b^i(z)$ are the redshift distribution and the redshift-dependent bias of galaxies in sample i.

In this work we replace $b^i(z)$ with a redshift-independent fiducial linear bias for each tomographic bin and only use clustering auto-correlations (i.e., $i = j$) as cosmological signal. All power spectra are calculated using PyCCL [52].

5.2.2 Cosmic shear

Galaxy clustering is a powerful probe of cosmology, but galaxies are biased tracers of the matter density field. Without a priori knowledge of the galaxy bias for a given sample, cosmological constraints are significantly degraded. Combining clustering with a tracer that directly probes the matter field, therefore, greatly enhances cosmological constraining power. Cosmic shear is such a probe.

---

[2]Note by assuming the shot noise is position-independent we have discarded a term $\delta_{ij}\,\delta(\boldsymbol{\theta})/\bar{n}_g$.

Photons emitted by galaxies are deflected by gravitational potentials associated with the large-scale structure of the universe. This effect, known as gravitational lensing, shears images of distant galaxies, so that circular galaxies appear as ellipses. For the majority of sources this shear is very weak. By measuring correlations of galaxy shapes across large scales, however, the matter distribution of the universe can be inferred.

The observed ellipticity of a galaxy can be written

$$\hat{\epsilon}(\boldsymbol{\theta}) = \gamma_G(\boldsymbol{\theta}) + \epsilon_I(\boldsymbol{\theta}) + \epsilon_{\rm rnd}, \tag{5.11}$$

where γ_G is the distortion of the ellipticity by weak lensing shear, ϵ_I is the intrinsic ellipticity induced by the local environment (e.g. tidal forces) in which the galaxy resides, and $\epsilon_{\rm rnd}$ is a random, uncorrelated component that accounts for stochasticity in the intrinsic ellipticities of galaxies. The two-point correlation function between the shapes of galaxies in two sample i and j is

$$\langle\hat{\epsilon}^i(\boldsymbol{\theta})\hat{\epsilon}^j(\boldsymbol{\theta})\rangle = \langle\gamma^i_G(\boldsymbol{\theta})\gamma^j_G(\boldsymbol{\theta})\rangle + \langle\gamma^i_G(\boldsymbol{\theta})\epsilon^j_I(\boldsymbol{\theta})\rangle + \langle\epsilon^i_I(\boldsymbol{\theta})\gamma^j_G(\boldsymbol{\theta})\rangle + \langle\epsilon^i_I(\boldsymbol{\theta})\epsilon^j_I(\boldsymbol{\theta})\rangle + \langle\epsilon^i_{\rm rnd}\epsilon^j_{\rm rnd}\rangle, \tag{5.12}$$

and taking the Fourier transform, we have the sum of power spectra

$$\hat{C}^{ij}_{\epsilon\epsilon}(\ell) = C^{ij}_{GG}(\ell) + C^{ij}_{IG}(\ell) + C^{ij}_{GI}(\ell) + C^{ij}_{II}(\ell) + \sigma^2_\epsilon\,\delta_{ij}, \tag{5.13}$$

where we have defined $\sigma^2_\epsilon \equiv \langle\epsilon^i_{\rm rnd}\epsilon^j_{\rm rnd}\rangle$. These 2D power spectra can also be calculated as projections of the 3D matter power spectrum:

$$C^{ij}_{GG}(\ell) = \int {\rm d}\chi\, \frac{q^i_G(\chi)\, q^j_G(\chi)}{\chi^2} P_{\delta\delta}\left(k = \frac{\ell+\frac{1}{2}}{\chi}, z\right), \tag{5.14}$$

$$C^{ij}_{IG}(\ell) = \int {\rm d}\chi\, \frac{q^i_I(\chi)\, q^j_G(\chi)}{\chi^2} P_{\delta\delta}\left(k = \frac{\ell+\frac{1}{2}}{\chi}, z\right), \tag{5.15}$$

$$C^{ij}_{II}(\ell) = \int {\rm d}\chi\, \frac{q^i_I(\chi)\, q^j_I(\chi)}{\chi^2} P_{\delta\delta}\left(k = \frac{\ell+\frac{1}{2}}{\chi}, z\right), \tag{5.16}$$

with the transfer functions [158, 213, 270]

$$q^i_G(\chi) = (1+z)\frac{3\Omega_m H_0^2}{2c^2}\int_\chi^\infty {\rm d}\chi' n^i(z')\frac{dz'}{d\chi'}\frac{\chi(\chi'-\chi)}{\chi'} \tag{5.17}$$

$$q^i_I(\chi) = -A_{IA}C_1\rho_c\frac{\Omega_m}{D(z)}, \tag{5.18}$$

where again we have implicitly defined $z \equiv z(\chi)$ as the cosmological redshift at distance χ. This latter transfer function represents the Non-linear Alignment (NLA; [169]) model of intrinsic alignments which quantifies the intrinsic alignment of galaxies in terms of the strength of the local tidal field. C_1 is a normalization constant, ρ_c is the critical density of the universe today, and $D(z)$ is the linear growth factor.

The cosmological signal we are interested in here is the correlation of gravitational shears, C_{GG}^{ij}, which is an unbiased tracer of the matter density. We see, however, this signal is contaminated via three terms related to the intrinsic alignment of galaxies: C_{IG}^{ij}, C_{GI}^{ij}, C_{II}^{ij}. The first two terms encode the correlation between the intrinsic alignment of galaxies in one bin with the shear seen in another bin. This term arises because the large scale structure generating the tidal field that locally aligns galaxies in bin i also shears the images of sources in the background bin j. From this we can see that when $i \neq j$ (and when bins i and j do not overlap) one of these terms is necessarily zero, as the structure tidally aligning a background bin does not shear the images of galaxies in a foreground bin.

Cosmic shear has a very low signal-to-noise ratio (SNR) for an individual galaxy. Typically, $\gamma \sim 10^{-2}$, while $\sigma_\epsilon \sim 1$, yielding an SNR of order 1%. On top of this, intrinsic alignment correlations often contribute at the 30% level [169]. Cosmic shear surveys, therefore, measure the correlation in shapes of millions (or even billions) of galaxies, and must model and marginalize over the effect of intrinsic alignment. This enables the measurement of the matter power spectrum $P_{\delta\delta}$ via the combination of Equations 5.13 and 5.16.

5.2.3 Galaxy-galaxy lensing

In addition to measuring the autocorrelation of galaxy number densities and shapes, we can measure the cross-correlation between these quantities. This new observable, galaxy-galaxy lensing, measures the correlation between the shear measured on background (source) galaxies with positions of foreground (lens) galaxies, probing the large-scale matter distribution around the foreground galaxies. Galaxy-galaxy lensing also provides a link that connects the clustering and shear signals, and carries unique dependencies on the systematics that plague

those signals. The combination of the three (called "3×2pt" due to the use of three 2-point correlation functions) provides powerful leverage for simultaneously constraining cosmology and systematic errors.

The galaxy-galaxy lensing signal can be calculated by correlating Equations 5.4 and 5.11:

$$\langle \hat{\epsilon}^i(\boldsymbol{\theta}) \hat{n}^j(\boldsymbol{\theta}) \rangle = \langle \gamma_G^i(\boldsymbol{\theta}) n_g^j(\boldsymbol{\theta}) \rangle + \langle \epsilon_I^i(\boldsymbol{\theta}) n_g^j(\boldsymbol{\theta}) \rangle, \tag{5.19}$$

with the corresponding power spectra

$$C_{Gg}^{ij}(\ell) = \int \mathrm{d}\chi \, \frac{q_G^i(\chi)\, q_g^j(\chi)}{\chi^2} P_{\delta\delta}\left(k = \frac{\ell + \frac{1}{2}}{\chi}, z\right) \tag{5.20}$$

$$C_{Ig}^{ij}(\ell) = \int \mathrm{d}z \, \frac{q_I^i(\chi)\, q_g^j(\chi)}{\chi^2} P_{\delta\delta}\left(k = \frac{\ell + \frac{1}{2}}{\chi}, z\right). \tag{5.21}$$

Note that $\langle \epsilon_I(\boldsymbol{\theta}) n_g(\boldsymbol{\theta}) \rangle = 0$ when $i > j$, assuming bins i and j do not overlap (correlations for $i < j$ are generated by magnification bias; see Section 5.3).

By inspecting the power spectra and transfer functions for each probe, one can see the different dependence each probe carries on galaxy bias and intrinsic alignment. For example, considering galaxy bias,

$$\text{clustering}: \quad P_{gg} \sim b^2 P_{\delta\delta} \tag{5.22}$$

$$\text{galaxy-galaxy lensing}: \quad P_{Gg} \sim b P_{\delta\delta} \tag{5.23}$$

$$\text{cosmic shear}: \quad P_{GG} \sim P_{\delta\delta}. \tag{5.24}$$

The combination of these three probes efficiently breaks galaxy-bias degeneracies, providing significantly improved constraints on both systematics and cosmology.

5.2.4 CMB lensing

In addition to shearing images of background galaxies, the gravitational potential sourced by large scale structure also distorts our image of the CMB. Measurements of this lensing, provided by surveys such as Planck [233], enable complementary constraints from gravitational lensing. In particular, CMB lensing is subject to different systematic errors than galaxy

lensing measurements, and also provides a single source plane at a very high, precisely constrained redshift. This latter feature enables lensing constraints at higher redshifts than are possible using galaxy lensing measurements, which are difficult to perform on high-redshift galaxies.

The signal from cross-correlation of galaxies with CMB lensing is very similar to the signal from galaxy-galaxy lensing (Equation 5.20), except there is no contamination from intrinsic alignments, and the source galaxy redshift distribution is replaced with a single source at redshift $z^* = 1100$. In particular,

$$C^i_{\kappa g}(\ell) = \int d\chi \frac{q^i_g(\chi) q_\kappa(\chi)}{\chi^2} P_{\delta\delta}\left(k = \frac{\ell + \frac{1}{2}}{\chi}, z\right), \tag{5.25}$$

where

$$q_\kappa(\chi) = (1+z)\frac{3\Omega_m H_0^2}{2c^2}\frac{\chi(\chi^* - \chi)}{\chi^*}, \tag{5.26}$$

and $\chi^* \equiv \chi(z^*)$ is the comoving distance to the surface of last scattering.

For the forecasts in this chapter we use forecasts for CMB lensing as detected by the Simons Observatory [4].

5.3 Systematic Errors for LBGs

In addition to the cosmology signals described above, in this section we describe several sources of systematic error that we simulate for our LBG samples. We model these effects and marginalize over the corresponding parameters, which reduces the constraining power of our cosmological probes. We do not, however, attempt to model any biases that may arise due to incorrect modeling of these effects. Note many of these systematics also contaminate the signal and complicate the interpretation of low-redshift observables, however in this chapter we only model and interpret their impact on high-redshift galaxies. Systematics associated with CMB lensing are also important, but we leave these to future work.

5.3.1 Survey non-uniformity

Given a true number of LBGs in a patch of the sky, only a fraction are detected and selected for our sample. This fraction is a function of the survey depth. Assembling a magnitude-limited sample significantly reduces this effect, but there are still second-order effects that contaminate the signal. Even at a fixed magnitude above the survey limit, the depth of different survey pixels modulates the selection efficiency, e.g. via color cuts or photo-z estimation.

We can model this effect by modulating the number density via the contamination function $f(\boldsymbol{\theta})$:

$$\hat{n}_g(\boldsymbol{\theta}) = (1 + f(\boldsymbol{\theta}))\, n_g(\boldsymbol{\theta}) + n_{\rm rnd}(\boldsymbol{\theta}). \tag{5.27}$$

Again transforming to the over-density, we have

$$\hat{\delta}_g(\boldsymbol{\theta}) = \delta_g(\boldsymbol{\theta}) + f(\boldsymbol{\theta}) + f(\boldsymbol{\theta})\delta_g(\boldsymbol{\theta}) + \frac{n_{\rm rnd}(\boldsymbol{\theta})}{\bar{n}_g}. \tag{5.28}$$

We will assume the second-order term $f(\boldsymbol{\theta})\delta_g(\boldsymbol{\theta})$ is sub-dominant. As a result, the multiplicative modulation of n_g results in an additive modulation of δ_g. This modifies our equation for the observed galaxy-clustering power spectrum:

$$\hat{C}^{ij}_{gg}(\ell) = C^{ij}_{gg}(\ell) + C^{ij}_{ff}(\boldsymbol{\theta}) + \bar{n}_g^{-1}\, \delta_{ij} \tag{5.29}$$

where C^{ij}_{ff} is the power spectrum of the contaminating field $f(\boldsymbol{\theta})$. The auto-spectra ($i = j$) will directly contaminate our clustering signal, while the cross-spectra ($i \neq j$) will increase the covariance between the observed clustering in different tomographic bins.

We model the field $f(\boldsymbol{\theta})$ using the LSST survey simulations and LBG number density models of the previous chapter. This is described in more detail in Section 5.5, where we optimize our cuts to maximize the SNR of our signal.

Note some fraction of contamination from non-uniformity can be corrected via regression on maps of survey properties [23, 308]. To account for this effect, we rescale $C_{ff} \to a_f \cdot C_{ff}$. It is reasonable to assume this process can correct for $\sim 90\%$ of this effect, so we use a fiducial value of $a_f = 0.1$. Note a_f is fixed in our forecasts, not marginalized over.

5.3.2 Magnification bias

Not only does weak lensing induce shear in galaxy shapes, but it also *magnifies* galaxies. Magnification alters the number density of galaxies above a flux limit $f_{\rm lim}$ by

$$\hat{n}_g(\boldsymbol{\theta}; > f_{\rm lim}) = \frac{1}{\mu(\boldsymbol{\theta})} n_g(\boldsymbol{\theta}; > \frac{f_{\rm lim}}{\mu(\boldsymbol{\theta})}), \tag{5.30}$$

where $\mu(\boldsymbol{\theta})$ is the magnification factor. The factor of μ^{-1} multiplying the flux limit accounts for the (de-)magnification of galaxies above the limit when $\mu > 1$ (< 1), which effectively decreases (increases) the flux limit of the survey. The additional prefactor of μ^{-1} accounts for the stretching (compression) of solid angles by the lensing.

If we assume the cumulative number density near the flux limit is well approximated by a power law,

$$n_g(\boldsymbol{\theta}; > f_{\rm lim}) = k f_{\rm lim}^{-\alpha}, \tag{5.31}$$

we can write

$$\hat{n}_g(\boldsymbol{\theta}; > f_{\rm lim}) = \frac{1}{\mu(\boldsymbol{\theta})} k \left(\frac{f_{\rm lim}}{\mu(\boldsymbol{\theta})}) \right) \tag{5.32}$$

$$= \mu^{\alpha-1} n_g(\boldsymbol{\theta}; > f_{\rm lim}) \tag{5.33}$$

$$= (1 + 2(\alpha - 1)\kappa)\, n_g(\boldsymbol{\theta}). \tag{5.34}$$

In the last equality we have used the weak lensing approximation $\mu \approx 1 + 2\kappa$, where $\kappa \ll 1$ is the lensing convergence. Including this multiplicative factor in our equation for the over-density, we have

$$\hat{\delta}_g(\boldsymbol{\theta}) = \delta_g(\boldsymbol{\theta}) + 2(\alpha - 1)\kappa + f(\boldsymbol{\theta}) + \frac{n_{\rm rnd}(\boldsymbol{\theta})}{\bar{n}_g}, \tag{5.35}$$

which adds two additional terms to the observed power spectrum

$$\hat{C}^{ij}_{gg}(\ell) = C^{ij}_{gg}(\ell) + 4(\alpha - 1)^2 C^{ij}_{GG} + 2(\alpha - 1) C^{ij}_{Gg} + C^{ij}_{ff}(\boldsymbol{\theta}) + \bar{n}_g^{-1}\, \delta_{ij}. \tag{5.36}$$

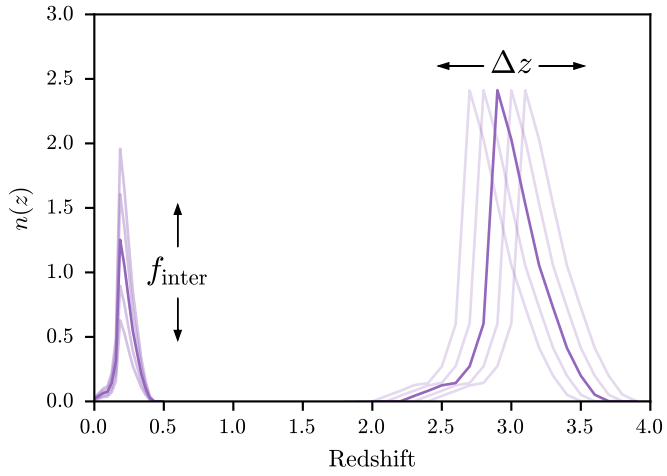

Figure 5.1: Schematic of the photo-z systematic error model, using the u-band dropouts as an example. Δz represents a shift of the true LBG redshift distribution; $f_{\rm inter}$ is the fraction of low-redshift interlopers resulting from confusion of the Lyman- and Balmer-breaks. Note this visualization omits the $(1 - f_{\rm inter})$ scaling of the true high-redshift LBG population.

Here we have used the fact that the convergence power spectrum is identical to the shear power spectrum[3]. The magnification auto-spectrum biases our measurements of galaxy clustering, while the cross term (which is only non-zero for $i < j$) will increase the covariance between the observed clustering in different tomographic bins.

We model magnification (i.e. estimate α) in our LBG samples by taking the derivative of the predicted number densities with respect to the magnitude cut, using the number density models from the previous chapter. We assume a Gaussian prior with $\sigma = 0.02$ for these parameters, matching the uncertainty in power-law models of the LBG luminosity function fit in the HSC GOLDRUSH survey [129].

5.3.3 Photo-z errors

Photo-z estimation is a key component of large scale structure surveys. They are needed to estimate $n^i(z)$ which is required to calculate all of the power spectra in the preceding sections. Photo-z estimation is notoriously difficult, however, and is a large source of systematic error in cosmology analyses with photometric galaxy surveys.

Cosmology from large scale structure is most sensitive to the mean redshift of each galaxy sample, and miscalibration of this value can seriously bias cosmological inference [218]. To mitigate this bias, previous surveys have introduced a Δz parameter that shifts the redshift distribution of a given sample (e.g, B. Hoyle et al. [137]). This parameter is then marginalized while performing inference on cosmological parameters. Marginalizing over these additional parameters decreases the bias of cosmological inference at the price of increasing the variance. We use a similar model, introducing a photo-z shift parameter for each tomographic bin: $n^i(z) \to n^i(z - \Delta z^i)$.

LBG selection also suffers from confusion of the Lyman and Balmer breaks, the latter of which is at $\sim 4000\,\text{Å}$. This confusion introduces a low-redshift interloper population into LBG samples, with a characteristic redshift corresponding to ratio between the wavelengths of the Lyman and Balmer breaks:

$$z_{\text{inter}} \sim \frac{\lambda_{\text{Balmer}}}{\lambda_{\text{Ly}\alpha}}(1+z) - 1. \tag{5.37}$$

We model interloper distribution by rescaling each LBG redshift distribution according to this ratio, and parameterize the fraction of interlopers with the parameter f_{inter}.

Combining the interloper and shift models, our model for the observed redshift distribution of each LBG sample is

$$\hat{n}^i(z) = f_{\text{inter}}\, n^i\left(\frac{\lambda_{\text{Ly}\alpha}}{\lambda_{\text{B}}}(1+z) - 1\right) + (1 - f_{\text{inter}})\, n^i(z - \Delta z^i). \tag{5.38}$$

We set fiducial values of $f_{\text{inter}} = 0.26, 0.05, 0.05$ in year 1 and $f_{\text{inter}} = 0.14, 0.02, 0.02$ by approximating the values found in [226], interpolated to LSST depths. We set Gaussian

[3]Note for the rest of this chapter we will reserve κ to refer to the convergence of CMB lensing

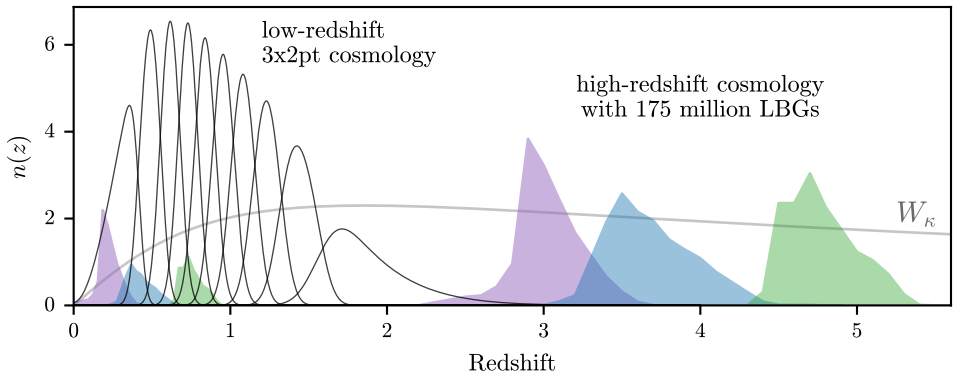

Figure 5.2: LBG redshift distributions assuming fiducial nuisance parameters. u-, g-, r-dropouts correspond to the purple, blue, green distributions, respectively. The black low-redshift redshift distributions are the redshift distributions assumed for the low-redshift 3x2pt analysis. The CMB lensing kernel is drawn in gray (labeled W_κ).

priors on these parameters with $\sigma = 0.05$. For the Δz^i parameters we set the fiducial value to 0. For the year 10 u-band dropout sample, we set a Gaussian prior of 0.01, assuming this parameter will be calibrated via cross-correlation with DESI-II spectroscopic tracers [266]. For all other parameters we do not set a prior, allowing this value to be entirely self-calibrated by the 3x2pt data.

5.4 Estimating signal covariance

To perform Fisher forecasts (see Section 5.1) we must estimate the covariance matrix of our signals. We construct two data vectors. The low-redshift data vector consists of galaxy clustering and cosmic shear auto-correlations, as well as galaxy-galaxy lensing cross-correlations between foreground lens samples and background source samples:

$$d_{\text{low-}z} = \bigoplus_{i=1}^{5\,(10)} \left\{ [\hat{C}^{ii}_{gg},\, \hat{C}^{ii}_{\epsilon\epsilon}] \bigoplus_{j<i} \hat{C}^{ij}_{\epsilon g} \right\}, \tag{5.39}$$

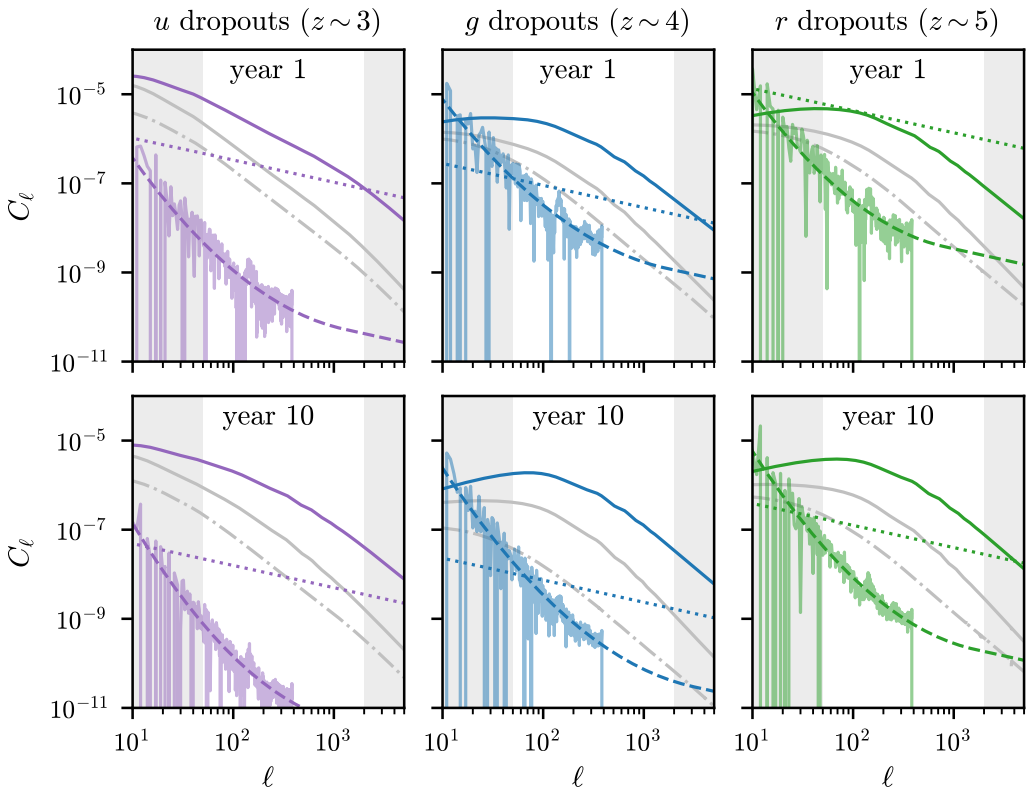

Figure 5.3: Galaxy clustering auto-spectra (solid) plotted against the noise terms in Equation 5.42 for LSST years 1 and 10. The solid gray lines in each panel represent cosmic variance; the dash-dotted gray lines represent magnification bias; the dotted lines represent shot noise; and the dashed lines represent the non-uniformity, for which the power-law models are plotted over the measured values (see Section 5.5). We see that, in year 1, u- and g-band dropouts are cosmic variance limited at large scales but shot noise limited on small scales, while r-band dropouts are shot noise limited on all scales. By year 10, however, u- and g-band dropouts are cosmic variance limited on all scales of interest, while r-band dropouts are shot-noise limited only at the smallest scales.

where i, j run over the 5 (10) tomographic bins assumed for LSST year 1 (10). The high-redshift data vector consists of LBG clustering, cross-correlation with CMB lensing, and the CMB lensing auto-spectrum:

$$d_{\text{high-}z} = \bigoplus_{i=1}^{3} [\hat{C}^{ii}_{gg}, \hat{C}^{i}_{\kappa g}] \oplus \hat{C}_{\kappa\kappa}, \tag{5.40}$$

where i runs over the 3 LBG tomographic bins (corresponding to u, g, r dropouts).

We calculate Gaussian covariances[4] for these signals:

$$\text{Cov}\left[\hat{C}^{ij}_{AB}, \hat{C}^{kl}_{XY}\right] = \frac{1}{(2\ell+1) f_{\text{sky}}} \left\{\hat{C}^{ik}_{AX}\hat{C}^{jl}_{BY} + \hat{C}^{il}_{AY}\hat{C}^{jk}_{BX}\right\}. \tag{5.41}$$

For example, the variance of the clustering signal in one of the tomographic LBG bins is

$$\text{Var}\left[\hat{C}^{ii}_{gg}\right] = \frac{2}{(2\ell+1) f_{\text{sky}}} \left(C^{ij}_{gg}(\ell) + 4(\alpha-1)^2 C^{ij}_{GG} + C^{ij}_{ff}(\boldsymbol{\theta}) + \bar{n}_g^{-1}\,\delta_{ij}\right)^2, \tag{5.42}$$

which receives contributions from cosmic variance, magnification bias, non-uniformity contamination, and shot noise. These noise components are plotted in Figure 5.3 along with the large-scale structure signal C^{ii}_{gg}. In year 1, you can see the u- and g-dropout clustering signals are cosmic-variance limited at $\ell \lesssim 200$ and shot-noise limited above this. In year 1 r-band dropouts are severely shot-noise limited at all scales, so this sample can likely be safely ignored for large-scale structure studies in year 1 of LSST. By year 10, however, the u- and g-dropout clustering signals are cosmic-variance limited at all relevant scales ($50 \leq \ell \leq 2000$), while the r-band dropouts are cosmic-variance limited at $\ell \lesssim 400$. This shows the power of the massive volume probed by LSST. Like all LSST cosmology probes, the power of LBG cosmology will be limited by our ability to constrain and model systematic errors, especially by year 10 of the survey.

The covariance for the high-redshift data vector is shown in Figure 5.4. The matrix is divided into blocks corresponding to u-, g-, and r-band dropouts (color coded purple, blue, green, respectively), each of which is further divided into sub-blocks corresponding to

[4]The Gaussian covariance is diagonal in ℓ, but the survey mask generates mode mixing, adding off-diagonal $\ell \neq \ell'$ power to the covariance matrix. We model this mixing using NaMaster [11].

galaxy clustering and cross-correlation with CMB lensing. There is an additional final block corresponding to the auto-correlation of CMB lensing (color coded black). The rows/columns within each sub-block correspond to 20 logarithmic bins between $50 \leq \ell \leq 2000$. Each sub-block is nearly diagonal, except for the off-diagonal large-scale power generated by the survey mask.

Note we model the covariance matrix for the high- and low-redshift data vectors separately, and do not account for the covariance between these two data sets. This underestimates the degradation in constraints when considering the impact of magnification bias, as the low-redshift structure will be correlated with the magnification bias of the high-redshift galaxies. We do not believe this omission substantially alters the conclusions of this chapter, however, and leave the modeling of this high-to-low covariance to future work.

5.5 Optimizing LBG selection

As noted in the Section 4.3.3, when assembling LBG samples we have two primary magnitude cuts we must select: the first is the magnitude limit for selecting LBGs; the second is the minimum depth for sky pixels that are included in our footprint. Deep LBG magnitude limits allows for higher number densities, but deeper limits can only be accommodated by sufficiently deep sky pixels. Reducing the magnitude limit reduces the average number density, but increases the fraction of the sky we can include in our survey. Determining the minimum depth for pixels in our survey has a similar effect, while also modulating the efficiency with which we can select LBGs. Setting the required depth independently of the LBG magnitude cut effectively determines what is the minimum SNR of LBG detection we will allow into our sample.

Allowing for a wider survey area also increases the amount of survey non-uniformity, as we progressively add pixels with a wider range of depths. This contaminates our clustering signal as described in Section 5.3.1.

For a given pair of values, $m_{\rm cut}$ and $m_{5,\rm min}$, one can calculate the clustering SNR by selecting the pixels that pass these cuts in the OpSim depth maps. This enables a simple

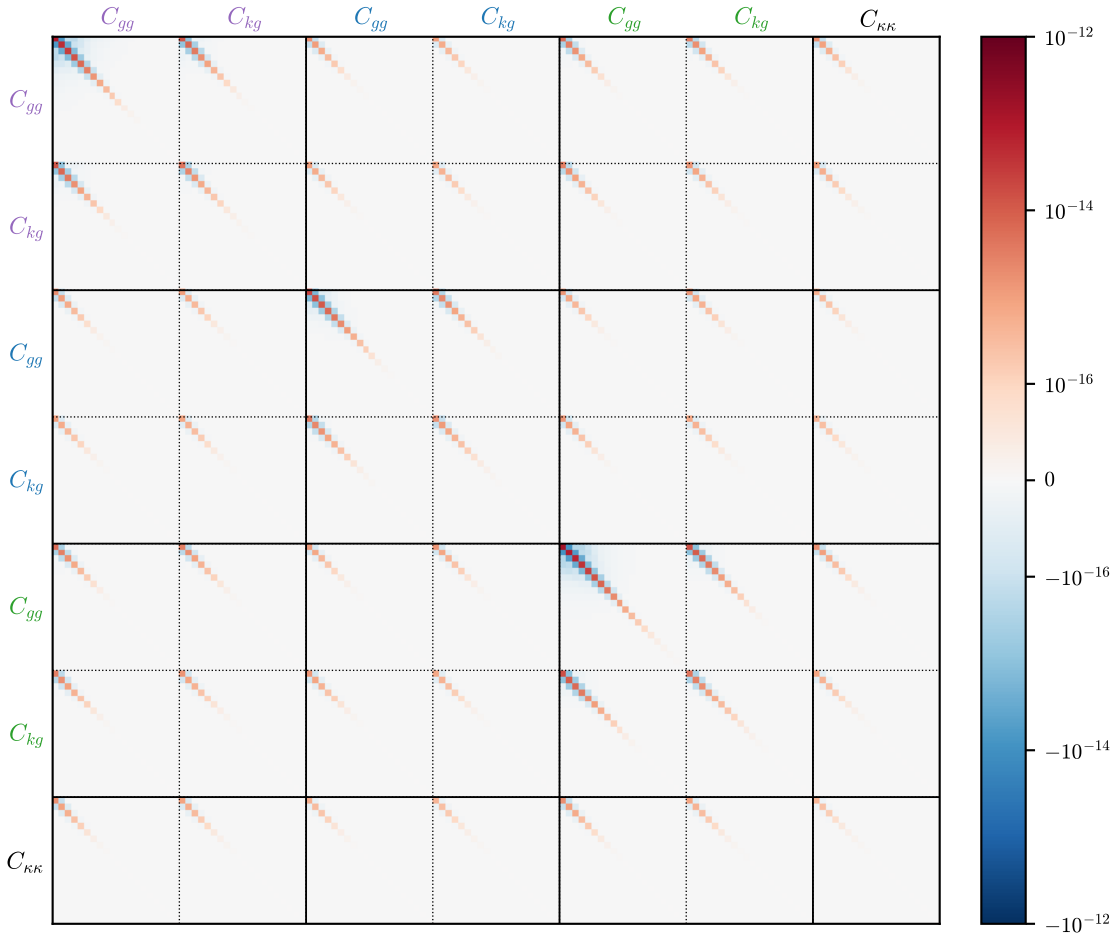

Figure 5.4: Covariance of LBG signal forecast for year 10 of LSST. The matrix is divided into blocks corresponding to u-, g-, and r-band dropouts (color coded purple, blue, green, respectively). The final block corresponds to the CMB lensing auto-correlation. Rows/columns within each block correspond to 20 logarithmic bins between $50 \leq \ell \leq 2000$. Off-diagonal covariance within each block is the result of mode coupling due to the survey mask.

calculation of $f_{\rm sky}$, and using the number density models from the previous chapter, one can calculate the LBG number density per OpSim pixel. This number density map can then be converted to an over-density map, where the over-densities are only due to depth variations, not large scale structure. Taking the power spectrum of this over-density map yields C_{ff}, and $\mathrm{Var}[\hat{C}_{gg}]$ is calculated using Equation 5.42. The SNR of clustering is then

$$\mathrm{SNR} = \sqrt{\sum_{\ell} \frac{\hat{C}_{gg}(\ell)}{\mathrm{Var}[\hat{C}_{gg}(\ell)]}}. \tag{5.43}$$

To balance the trade-off between survey area and number density we optimize $m_{\rm cut}$ and $m_{5,\rm min}$ to maximize this SNR. Note we constrain the optimization such that $f_{\rm sky} > 0.25$ to prevent the survey footprint from becoming overly shredded.

The results of this optimization are shown in Figure 5.5. You can see the survey uniformity dramatically improves by year 10 of LSST, allowing the use of nearly the entire Wide Fast Deep survey footprint. By contrast, year 1 is significantly less uniform, necessitating the removal of a significant fraction of the footprint due to insufficient survey depth. Quantitative results of these cuts are listed in Table 5.1. This table lists results for limiting values of $a_f = 0.0, 1.0$, in addition to the fiducial value of $a_f = 0.1$. The results are only weakly sensitive to a_f, however, indicating survey non-uniformity is not a major driver of the optimization. This conclusion agrees with the results shown in Figure 5.3 which shows the contamination from non-uniformity is strongly subdominant to other noise terms. In particular, the two main factors driving the optimization of the magnitude cuts are the shot noise and sky fraction.

5.6 Results

In this section we show results of forecasting cosmology constraints from the low- and high-redshift 3×2pt analyses individually, then the gain in constraining power by combining both constraints in a single analysis, while marginalizing over all nuisance parameters. We will focus on two pairs of parameters:

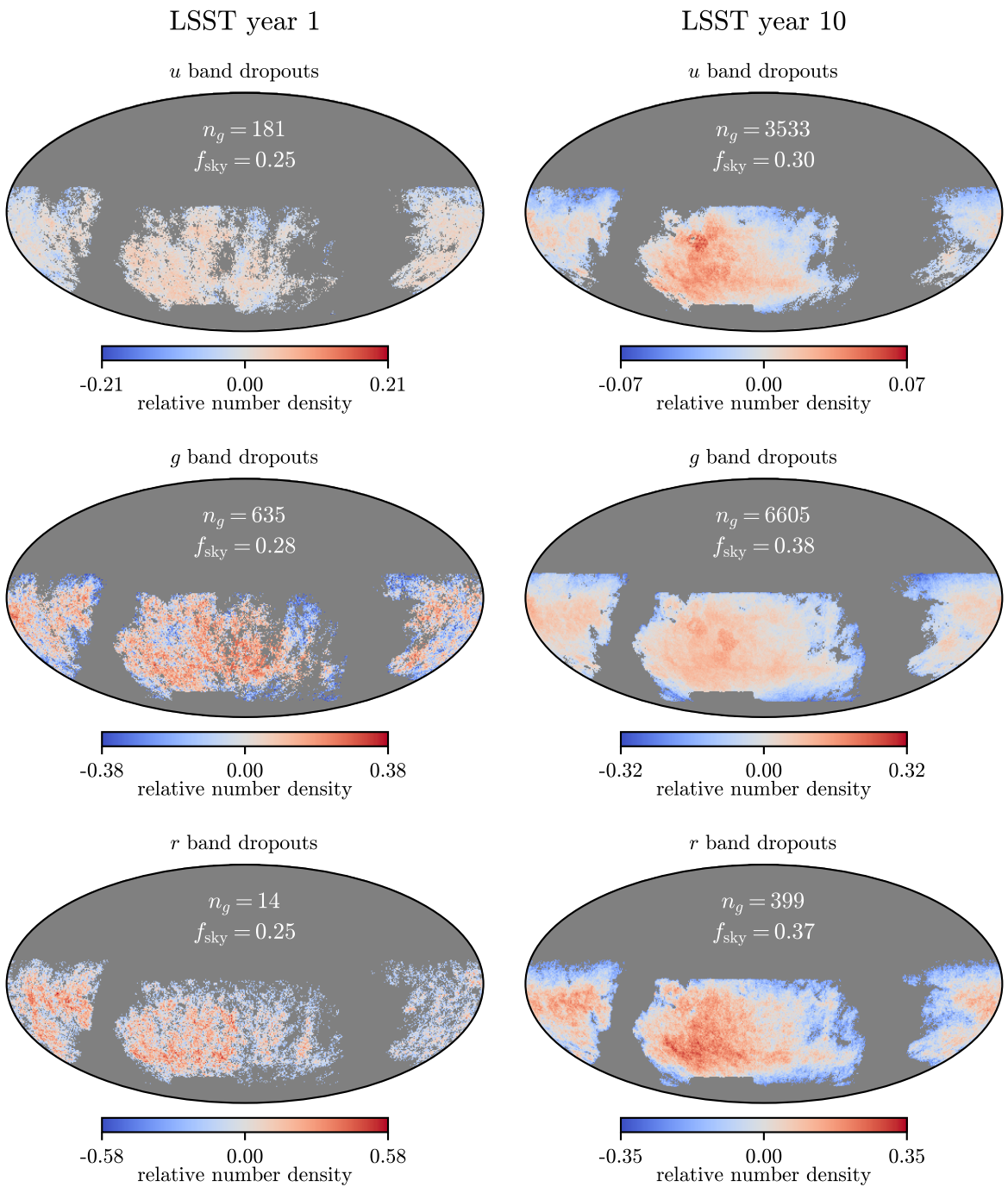

Figure 5.5: Over-density maps for LBG dropout samples, after optimizing the magnitude cuts. The corresponding average number densities and sky fractions are print in each panel.

u dropouts

	Year 1					Year 10				
a_f	$r_{\rm cut}$	$r_{5,\rm min}$	n_g	$f_{\rm sky}$	SNR	$r_{\rm cut}$	$r_{5,\rm min}$	n_g	$f_{\rm sky}$	SNR
0.0	23.68	23.68	181	0.25	95	24.76	24.85	3533	0.30	405
0.1	23.68	23.68	181	0.25	95	24.76	24.85	3533	0.30	405
1.0	23.68	23.68	181	0.25	94	24.75	24.85	3463	0.30	404

g dropouts

	Year 1					Year 10				
a_f	$i_{\rm cut}$	$i_{5,\rm min}$	n_g	$f_{\rm sky}$	SNR	$i_{\rm cut}$	$i_{5,\rm min}$	n_g	$f_{\rm sky}$	SNR
0.0	24.76	24.76	657	0.27	179	25.85	25.87	6737	0.38	559
0.1	24.74	24.75	635	0.28	171	25.83	25.89	6605	0.38	554
1.0	24.58	24.65	469	0.34	140	25.71	25.88	5801	0.40	528

r dropouts

	Year 1					Year 10				
a_f	$z_{\rm cut}$	$z_{5,\rm min}$	n_g	$f_{\rm sky}$	SNR	$z_{\rm cut}$	$z_{5,\rm min}$	n_g	$f_{\rm sky}$	SNR
0.0	24.34	24.34	14	0.25	16	25.51	25.51	405	0.36	226
0.1	24.34	24.34	14	0.25	16	25.50	25.50	399	0.37	225
1.0	24.34	24.34	14	0.25	15	25.50	25.50	399	0.37	214

Table 5.1: Optimized cuts for dropout galaxies. Cuts for each sample are specified for the detection band. Number densities are in units deg^{-2}; SNR is for galaxy clustering only.

- S_8 and Ω_m, which determine the amplitude of the power spectra;
- w_0 and w_a, which determine the evolution of dark energy.

Note while forecasting the pair S_8 and Ω_m, we fix w_0 and w_a to their standard ΛCDM values ($w_0 = -1$ and $w_a = 0$, corresponding to a cosmological constant).

Figure 5.6 shows fiducial constraints on S_8 vs Ω_m and w_a vs w_0, in which the only systematics we consider are galaxy bias and intrinsic alignment. Low-redshift constraints are shown in black, high-redshift constraints in gold, and joint constraints in blue. The numbers in each panel represent the figure of merit (i.e., inverse area of 1σ ellipses) relative to the low-redshift constraint.

We see in all cases the high-redshift constraints are weaker than the low-redshift constraints, but the combination of the two provides significant gains in constraining power. In particular, in year 1 the degeneracy direction of the high-redshift constraints in the S_8 vs Ω_m plane is rotated with respect to the degeneracy direction of the low-redshift constraints. The intersection of the two, therefore, yields substantial improvements, reducing uncertainty by a factor of 3.3. By year 10, the difference in degeneracy direction is reduced due to the higher-redshift reach of the year 10 "low redshift" sample, however the combination of the two constraints still yields a substantial improvement.

The impact on dark energy constraints is similar. The high-redshift constraints by themselves are weaker than the low-redshift constraints, but the combination yields significant improvement. There is a natural intuition for why the high-redshift constraints provide so much extra leverage: when attempting to constrain the long-term evolution of dark energy, it is beneficial to include data from higher redshift, providing a longer lever arm to constrain the evolution.

The power of the high-redshift dark energy constraints can be seen more clearly in the binned constraints of Figure 5.7. The back bars show the constraints from DESI [62], which are highly constraining at low-redshifts, but much weaker above $z = 1$. In year 1, LSST is expected to deliver constraints similar to the DESI high-redshift constraints, extending

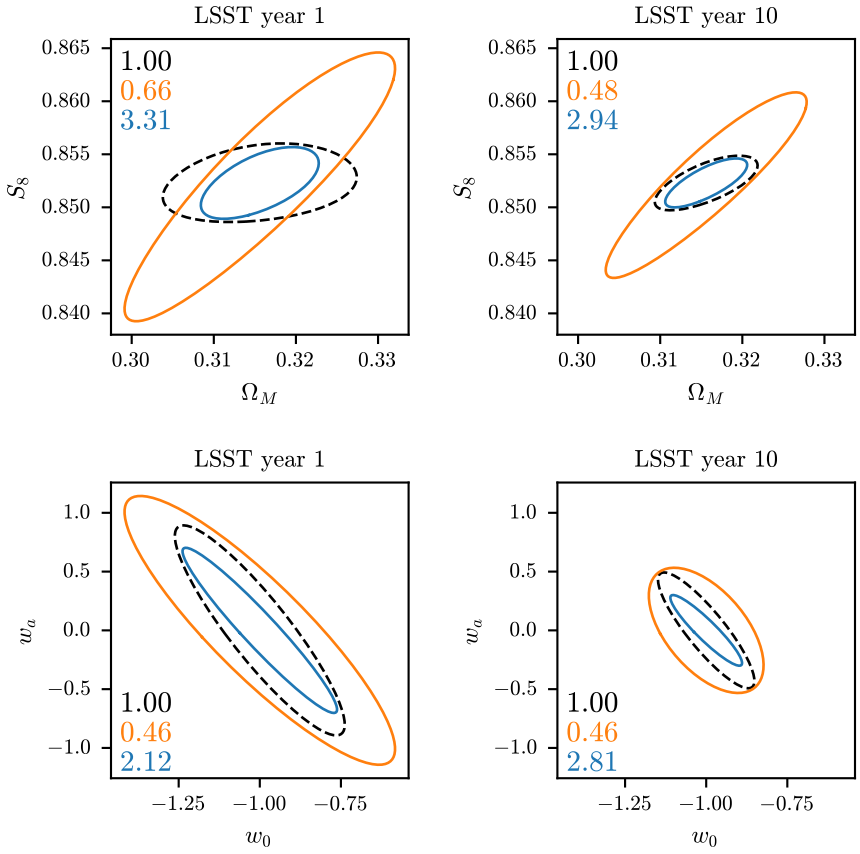

Figure 5.6: Constraints on cosmology using high (gold) and low (black) redshift 3×2pt analyses. The blue contours show the joint constraints. The numbers in each panel represent the figure of merit (i.e., inverse area of the ellipse) relative to the low-redshift constraint. The left column shows constraints for year 1; the right column shows constraints for year 10. The only systematic errors considered for these forecasts are the galaxy bias and intrinsic alignment. In all cases the high-redshift constraints are weaker in isolation, but yield significant improvements when combined with the low-redshift constraints.

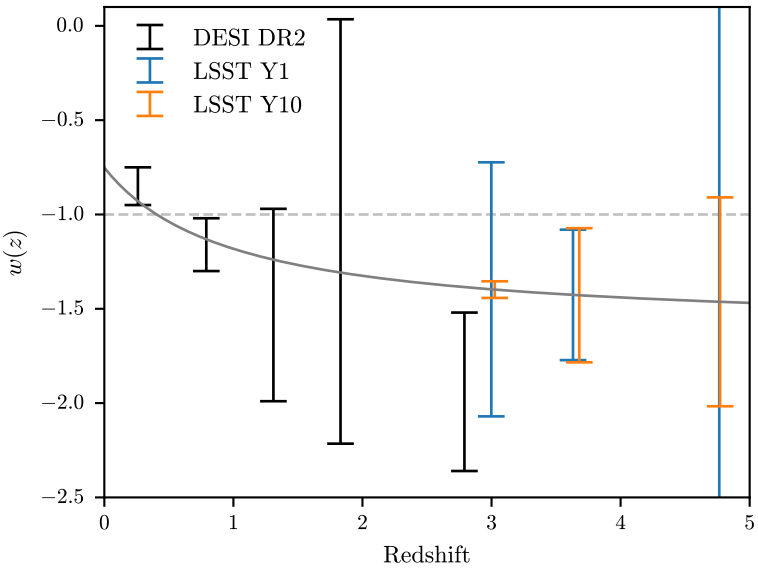

Figure 5.7: Binned constraints on the dark energy equation of state. The black bars are DESI constraints from [62]. The blue (gold) bars are forecasts for year 1 (10) LSST LBGs. The LSST LBGs significantly improve constraints at high redshift.

them to redshift $z \sim 4$. In year 10, however, LSST is expected to deliver extremely tight constraints at $z \sim 3$, and extend constraints to $z \sim 5$. The forecast LSST constraints for g-band dropouts ($z \sim 4$) do not improve between years 1 and 10, likely indicating this sample is largely systematics limited. More investigation on this point is needed.

Finally, Figure 5.8 shows the impact of systematic errors on the dark energy constraints[5]. This figure includes constraints using clustering only ($g \times g$), the cross-correlation of LBG positions with CMB lensing ($\kappa_{\rm CBM} \times g$), the combination of these two probes (2×2pt), and the full 3×2pt combination. This allows us to see how each probe is impacted by and constrains different systematic errors. For example:

- In general, while the clustering signal is more constraining than the CMB lensing cross-correlation, it is also more significantly impacted by marginalizing over systematic errors.

[5]While not shown here, similar conclusions can be reached for the S_8 vs Ω_m constraints

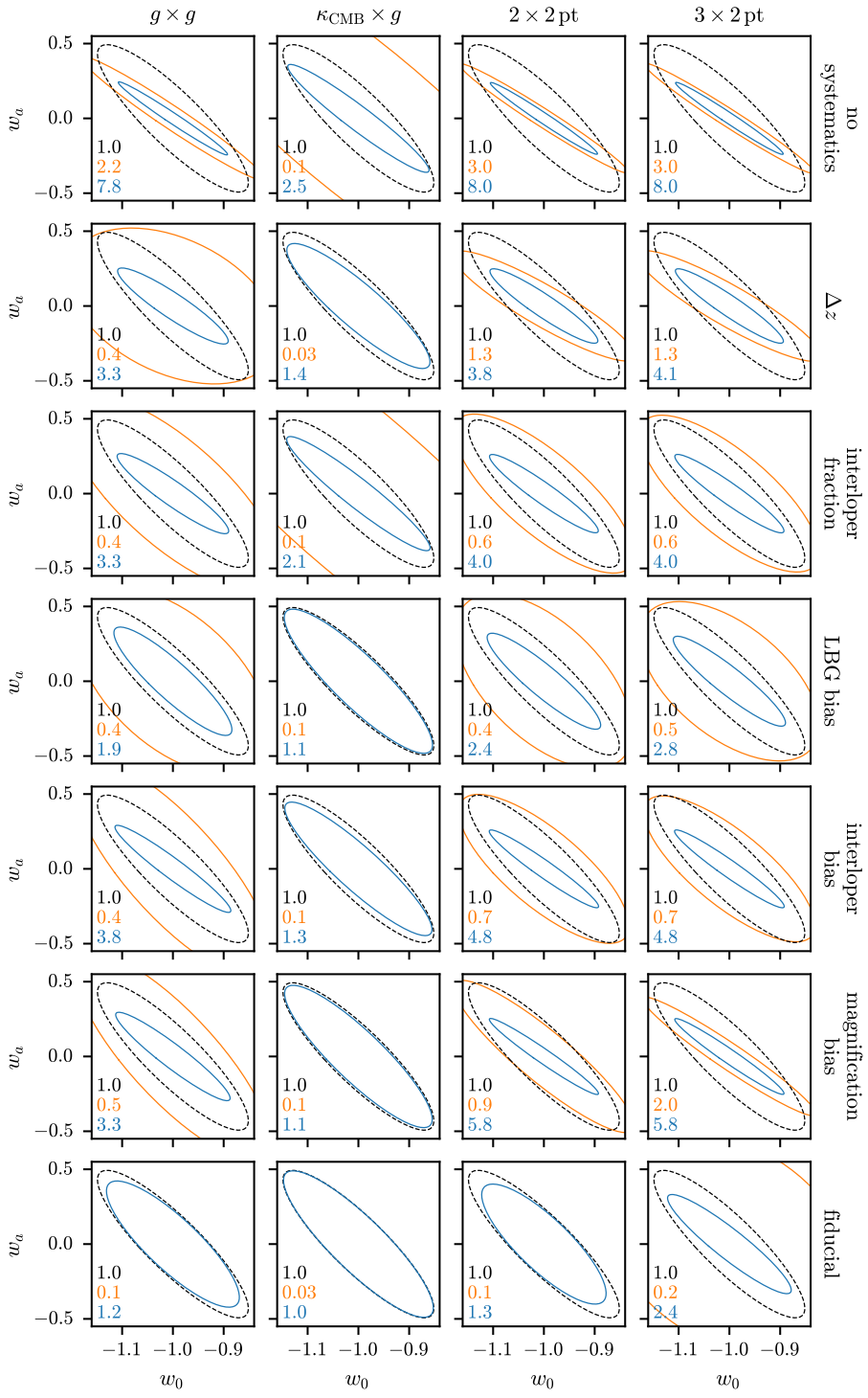

Figure 5.8: Caption on next page

Figure 5.8: (on preceding page) Comparison of LBG constraining power on the evolution of dark energy while marginalizing over different systematics errors. Each row corresponds to a different systematic error (label on right), while each column corresponds to a different data combination (labels on top).

Considering the LBG-only constraints, the degradation from marginalizing over Δz is approximately 40% worse for clustering than for the cross-correlation; from marginalizing over the interloper fraction, 400% worse; from marginalizing over the bias parameters, 150% worse; from marginalizing over all nuisance parameters together, 300% worse.

- the CMB lensing auto-spectrum provides no additional information in the absence of systematic errors, nor, in the LBG-only scenario, does it help constrain most systematic errors. It does, however, provide significant leverage to constrain magnification bias, as the CMB lensing auto-spectrum contains lensing from the same foreground structure that gives rise to the magnification bias.

- The CMB lensing auto-spectrum plays a different role in the joint high-/low-redshift scenario. In this case, magnification bias is already well constrained in the 2×2pt scenario, but the CMB lensing auto-spectrum provides additional power to constrain the Δz and LBG bias parameters. These effects are likely due to the changing levels of degeneracy between different parameters due to the altogether stronger constraints. Stronger conclusions will require a more detailed analysis.

- for the joint analysis under the fiducial model with all systematics marginalized, the 3×2pt combination provides significant improvement over the 2×2pt combination, which is only slightly more constraining than galaxy clustering alone.

Note all of these conclusions concern only the increase in variance on constrained parameters, and do not consider how these systematic errors may bias the cosmological constraints. When including the impact of bias, the importance and power of the 3×2pt combination will only be heightened, as each of the three probes is impacted by different systematic errors, and the combination of the three builds robustness to errors that impact only a subset of the three probes.

Recall that all of these systematic errors will also impact the low-redshift 3×2pt analysis. As we have only modeled these systematics for LBGs, Figure 5.8 is overly generous to the low-redshift constraints, and therefore undersells the value of the high-redshift constraints.

5.7 Conclusion

In this chapter we have forecast high-redshift cosmology constraints using the clustering of LBGs detected by LSST, and their cross-correlation with CMB lensing measured by the Simons Observatory. We combined these high-redshift observables into a high-redshift 3×2pt analysis, and considered the combination of these constraints with LSST low-redshift 3×2pt constraints using galaxy clustering and cosmic shear.

We find the high-redshift constraints, in isolation, are weaker than the low-redshift constraints, but that they are highly complementary to the low-redshift constraints, with the combination of the 2 yielding significant gains for cosmology. In particular, year 10 constraints on S_8 vs Ω_m and w_a vs w_0 are both improved by a factor of approximately 3.

We considered different combinations of the clustering, CMB lensing cross-correlation, and CMB lensing auto-correlation signals, concluding that the combination of all three significantly improves cosmological constraints while marginalizing over systematic errors models. We find that even when marginalizing over all systematic error models, the high-redshift 3×2pt analysis yields significant gains over low-redshift cosmology alone.

In the future we will extend these forecasts to other cosmological parameters, such as the sum of neutrino masses; we will include the full covariance between the low- and high-redshift tracers induced by magnification bias; we will also evaluate the impact of adding

more nuisance parameters associated with LBG galaxy bias to model scale-dependent galaxy bias [312].

Chapter 6

CONCLUSION

This decade is an extremely exciting era in which to study cosmology. We have achieved precision measurements of the key quantities that define our models, and progress now relies on ever more detailed understanding of our instruments, the signals we study, and the systematic errors that confound them. Despite these successes, our cosmological model remains mysteriously phenomenological at heart. We continue to test this model with more data and different probes, on different scales, in different epochs, *and the model is starting to break!*

We do not yet know if these are signs of new physics, or whether our modeling and understanding of our probes is breaking down. However this decade will see the fruition of a number of huge, transformational cosmology surveys. DESI is publishing provocative results, Euclid has released their first data, LSST is set to begin this fall, and Roman is scheduled to launch next year. Over the next few years, this deluge of data will certainly transform our understanding of the origin and evolution of our universe.

This thesis details my past and ongoing contributions to the project of better understanding our universe and pushing our understanding to higher redshifts and earlier times. Chapters 2 and 3 detailed efforts to better model and understand key systematic errors that must be conquered if we are to achieve the full potential of the photometric cosmology surveys coming this decade. The former developed an algorithm for learning galaxy spectral templates directly from broadband photometry, and demonstrated these templates outperform standard sets when applied to photo-z estimation. The latter developed new deep learning methodologies for simulating photometric galaxy catalogs with the goal of simulating and calibrating systematic errors in photo-z estimation. These two chapters derive from

my first-author publications contributing to the literature on photo-z calibration. During my PhD I was also a coauthor on the following related studies that applied the methods of Chapter 3 to various problems in cosmology:

- A. I. Malz et al. [195] quantifies the impact of survey strategy on photo-z estimation, using ideas from information theory;
- N. Stylianou et al. [286] studies the impact of non-representative training sets and training set errors on photo-z estimation;
- M. Lokken et al. [184] simulates correlated host galaxies for a synthetic catalog of supernovae and other transients;
- I. Moskowitz et al. [210] studies how simulations can be used to augment spectroscopic training sets to reduce the problem of non-representativity;
- Q. Hang et al. [127] studies how survey non-uniformity leads to anisotropies in photo-z distributions, and how these anisotropies propagate to biases in cosmological analyses.

Chapters 4 and 5 detailed efforts to optimize LSST for high-redshift science and forecast cosmology constraints. The former developed a data-driven model for LBG populations detected by LSST, including forecasts for detected number densities. The work of this chapter was submitted to the Rubin Observatory Survey Cadence Optimization Committee and factored in their decision to adapt the LSST survey strategy in a manner that improves high-redshift cosmology constraints. The latter chapter used the LBG population model to forecast constraints on S_8, Ω_m, and the evolution of dark energy, using the clustering of LBGs, their cross-correlation with CMB lensing, and their joint analyses with low-redshift clustering and cosmic shear measurements. This chapter demonstrated the power of LBG cosmology with LSST, forecasting a 3x increase in constraining power over that of low-redshift data alone. These results, including the analysis of systematic errors, will help guide our LBG cosmology efforts in the early years of LSST.

Outside the main body of the thesis, the appendices that follow explain my contributions to developing and commissioning the Rubin Observatory's active optics system (AOS). This work aims to improve the speed and robustness of the Rubin AOS, to ensure LSST delivers the image quality necessary to achieve precision science.

6.1 Future Directions

As I write this conclusion, Rubin-LSST is just beginning. Euclid has just released their first data, the Roman Space Telescope is scheduled to launch next year, and DESI-II is only a few years away. The coming decade and the data it brings will undoubtedly transform our view of the high-redshift universe, and may finally overturn the ΛCDM model. Achieving the full potential of this data will require the development of new analysis tools and detailed studies to understand the subtleties of this previously unprobed epoch.

Studies that I hope to pursue include:

- Adapting the template learning model of Chapter 2 to a deep learning paradigm, to incorporate more data, including photometric data without spectroscopic redshifts, then apply this model to learn spectral templates for high-redshift galaxies;

- Using photometric data, color distributions, and UV luminosity functions to improve the high-redshift calibration of cosmological simulations so they can be used for studying systematic errors and validating analysis choices for high-redshift cosmology;

- Exploring methods for photo-z calibration of high-redshift galaxies, including using photometry from the LSST, Euclid, and Roman deep fields, as well as multiband data from Hubble and JWST, and whatever limited spectroscopic information is available at high redshifts;

- Using cross-correlations with low-redshift spectroscopic tracers to characterize the redshift distribution and galaxy bias of low-redshift interloper populations that contaminate photometric LBG samples;

- Using low-redshift constraints from LSST and other surveys to remove the low-redshift contribution to CMB lensing, thereby boosting the cross-correlation signal with high-redshift LBGs, and to model the magnification bias of high-redshift galaxies;

- Studying how to model the scale-dependent galaxy bias observed in LBG populations [96, 312] to minimize the degradation of constraints while simultaneously avoiding bias in cosmological parameters.

These efforts will complement the explosion of interest in high-redshift galaxies spurred by JWST. As our understanding of high-redshift cosmology improves, we will better understand the environment in which these galaxies live and grow. Improving our understanding of these galaxies will, in turn, improve our understanding and control of systematic errors in high-redshift cosmology. Photometric analyses with LSST and spectroscopic analyses with DESI-II promise to provide an unprecedented view of large scale structure in the high-redshift universe, expanding the purview of precision cosmology, and generating a more comprehensive view of the cosmos.

BIBLIOGRAPHY

[1] Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing, PhRvD, 105, 023520, doi: 10.1103/PhysRevD.105.023520

[2] Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2023, Dark Energy Survey Year 3 results: Constraints on extensions to Λ CDM with weak lensing and galaxy clustering, PhRvD, 107, 083504, doi: 10.1103/PhysRevD.107.083504

[3] Adams, N. J., Bowler, R. A. A., Jarvis, M. J., et al. 2020, The rest-frame UV luminosity function at z 4: a significant contribution of AGNs to the bright end of the galaxy population, MNRAS, 494, 1771, doi: 10.1093/mnras/staa687

[4] Ade, P., Aguirre, J., Ahmed, Z., et al. 2019, The Simons Observatory: science goals and forecasts, JCAP, 2019, 056, doi: 10.1088/1475-7516/2019/02/056

[5] Adil, S. A., Mukhopadhyay, U., Sen, A. A., & Vagnozzi, S. 2023, Dark Energy in Light of the Early JWST Observations: Case for a Negative Cosmological Constant? Journal of Cosmology and Astroparticle Physics, 2023, 072, doi: 10.1088/1475-7516/2023/10/072

[6] Ahmad, Q. R., Allen, R. C., Andersen, T. C., et al. 2001, Measurement of the Rate of ν_e + d –¿ p + p + e^- Interactions Produced by ^{8}B Solar Neutrinos at the Sudbury Neutrino Observatory, PhRvL, 87, 071301, doi: 10.1103/PhysRevLett.87.071301

[7] Ahmad, Q. R., Allen, R. C., Andersen, T. C., et al. 2002, Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory, PhRvL, 89, 011301, doi: 10.1103/PhysRevLett.89.011301

[8] Ahn, M. H., Aliu, E., Andringa, S., et al. 2006, Measurement of neutrino oscillation by the K2K experiment, PhRvD, 74, 072003, doi: 10.1103/PhysRevD.74.072003

[9] Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, Second data release of the Hyper Suprime-Cam Subaru Strategic Program, PASJ, 71, 114, doi: 10.1093/pasj/psz103

[10] Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, Report of the Dark Energy Task Force, arXiv e-prints, astro, doi: 10.48550/arXiv.astro-ph/0609591

[11] Alonso, D., Sanchez, J., Slosar, A., & LSST Dark Energy Science Collaboration. 2019, A unified pseudo-C_ℓ framework, MNRAS, 484, 4127, doi: 10.1093/mnras/stz093

[12] Alvarez, M. A., Busha, M., Abel, T., & Wechsler, R. H. 2009, Connecting Reionization to the Local Universe, ApJL, 703, L167, doi: 10.1088/0004-637X/703/2/L167

[13] Amon, A., & Efstathiou, G. 2022, A Non-Linear Solution to the S_8 Tension? Monthly Notices of the Royal Astronomical Society, 516, 5355, doi: 10.1093/mnras/stac2429

[14] Angeli, G. Z., Xin, B., Claver, C., et al. 2014, in Modeling, Systems Engineering, and Project Management for Astronomy VI, Vol. 9150 (SPIE), 159–174, doi: 10.1117/12.2055390

[15] Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, Measuring and modelling the redshift evolution of clustering: The Hubble Deep Field North,, Tech. Rep. 2, doi: 10.1046/j.1365-8711.1999.02978.x

[16] Athar, M. S., Barwick, S. W., Brunner, T., et al. 2022, Status and perspectives of neutrino physics, Progress in Particle and Nuclear Physics, 124, 103947, doi: 10.1016/j.ppnp.2022.103947

[17] Beacom, J. F., Bell, N. F., & Dodelson, S. 2004, Neutrinoless Universe, PhRvL, 93, 121302, doi: 10.1103/PhysRevLett.93.121302

[18] Bechtol, K., Birrer, S., Cyr-Racine, F.-Y., et al. 2022, Snowmass2021 Cosmic Frontier White Paper: Dark Matter Physics from Halo Measurements, arXiv, doi: 10.48550/arXiv.2203.07354

[19] Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, The Hubble Ultra Deep Field, AJ, 132, 1729, doi: 10.1086/507302

[20] Bellomo, N., Bellini, E., Hu, B., et al. 2017, Hiding neutrino mass in modified gravity cosmologies, JCAP, 2017, 043, doi: 10.1088/1475-7516/2017/02/043

[21] Benítez, N. 2000, Bayesian Photometric Redshift Estimation, The Astrophysical Journal, 536, 571, doi: 10.1086/308947

[22] Benitez, N., Ford, H., Bouwens, R., et al. 2004, Faint Galaxies in Deep Advanced Camera for Surveys Observations, The Astrophysical Journal Supplement Series, 150, 1, doi: 10.1086/380120

[23] Berlfein, F., Mandelbaum, R., Dodelson, S., & Schafer, C. 2024, Joint Inference of Multiplicative and Additive Systematics in Galaxy Density Fluctuations and Clustering Measurements, Monthly Notices of the Royal Astronomical Society, 531, 4954, doi: 10.1093/mnras/stae1444

[24] Bessell, M. S. 2005, Standard Photometric Systems, Annual Review of Astronomy and Astrophysics, 43, 293, doi: 10.1146/annurev.astro.41.082801.100251

[25] Bianco, F. B., Ivezić, Ž., Jones, R. L., et al. 2022, Optimization of the Observing Cadence for the Rubin Observatory Legacy Survey of Space and Time: A Pioneering Process of Community-focused Experimental Design, ApJS, 258, 1, doi: 10.3847/1538-4365/ac3e72

[26] Bird, S., Albert, A., Dawson, W., et al. 2023, Snowmass2021 Cosmic Frontier White Paper: Primordial black hole dark matter, Physics of the Dark Universe, 41, 101231, doi: 10.1016/j.dark.2023.101231

[27] Bolte, M., & Hogan, C. J. 1995, Conflict over the age of the Universe, Nature, 376, 399, doi: 10.1038/376399a0

[28] Bousso, R. 2012, The Cosmological Constant Problem, Dark Energy, and the Landscape of String Theory, arXiv e-prints, arXiv:1203.0307, doi: 10.48550/arXiv.1203.0307

[29] Bouwens, R. J., Illingworth, G. D., van Dokkum, P. G., et al. 2022, Sizes of Lensed Lower-luminosity z = 4-8 Galaxies from the Hubble Frontier Field Program, ApJ, 927, 81, doi: 10.3847/1538-4357/ac4791

[30] Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, UV-continuum Slopes of >4000 z ˜ 4-8 Galaxies from the HUDF/XDF, HUDF09, ERS, CANDELS-South, and CANDELS-North Fields, The Astrophysical Journal, 793, 115, doi: 10.1088/0004-637X/793/2/115

[31] Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, New Determinations of the UV Luminosity Functions from z 9 to 2 Show a Remarkable Consistency with Halo Growth and a Constant Star Formation Efficiency, The Astronomical Journal, 162, 47, doi: 10.3847/1538-3881/abf83e

[32] Bowler, R. A. A., Jarvis, M. J., Dunlop, J. S., et al. 2020, A Lack of Evolution in the Very Bright End of the Galaxy Luminosity Function from z ≃ 8 to 10, Monthly Notices of the Royal Astronomical Society, 493, 2059, doi: 10.1093/mnras/staa313

[33] Boylan-Kolchin, M. 2022, Stress Testing ΛCDM with High-redshift Galaxy Candidates, https://ui.adsabs.harvard.edu/abs/2022arXiv220801611B

[34] Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, JAX: Composable Transformations of Python+NumPy Programs, http://github.com/google/jax

[35] Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, EAZY: A Fast, Public Photometric Redshift Code, The Astrophysical Journal, 686, 1503, doi: 10.1086/591786

[36] Brout, D., Scolnic, D., Popovic, B., et al. 2022, The Pantheon+ Analysis: Cosmological Constraints, ApJ, 938, 110, doi: 10.3847/1538-4357/ac8e04

[37] Bruzual, G., & Charlot, S. 2003, Stellar population synthesis at the resolution of 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000, doi: 10.1046/j.1365-8711.2003.06897.x

[38] Bruzual A., G., & Charlot, S. 1993, Spectral evolution of stellar populations using isochrone synthesis, The Astrophysical Journal, 405, 538, doi: 10.1086/172385

[39] Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., & Dickinson, M. 2000, Creating Spectral Templates from Multicolor Redshift Surveys, The Astronomical Journal, 120, 1588, doi: 10.1086/301514

[40] Burgess, C. P. 2013, The Cosmological Constant Problem: Why it's hard to get Dark Energy from Micro-physics, arXiv e-prints, arXiv:1309.4133, doi: 10.48550/arXiv.1309.4133

[41] Busso, G., Cacciari, C., Carrasco, J. M., et al. 2018, Gaia DR2 Documentation Chapter 5: Photometry,, Tech. rep. https://ui.adsabs.harvard.edu/abs/2018gdr2.reptE...5B

[42] Caldwell, R. R., Komp, W., Parker, L., & Vanzella, D. A. T. 2006, Sudden gravitational transition, PhRvD, 73, 023513, doi: 10.1103/PhysRevD.73.023513

[43] Carlsten, S. G., Strauss, M. A., Lupton, R. H., Meyers, J. E., & Miyazaki, S. 2018, Wavelength-Dependent PSFs and Their Impact on Weak Lensing Measurements, Monthly Notices of the Royal Astronomical Society, 479, 1491, doi: 10.1093/mnras/sty1636

[44] Carroll, S. M., de Felice, A., & Trodden, M. 2005, Can we be tricked into thinking that w is less than -1?, PhRvD, 71, 023525, doi: 10.1103/PhysRevD.71.023525

[45] Castellano, M., Fontana, A., Treu, T., et al. 2022, Early Results from GLASS-JWST. III: Galaxy Candidates at Z$\sim$9-15, doi: 10.48550/arXiv.2207.09436

[46] Celoria, M., & Matarrese, S. 2018, Primordial Non-Gaussianity, arXiv e-prints, arXiv:1812.08197, doi: 10.48550/arXiv.1812.08197

[47] Chakraborty, A., Chanda, P. K., Das, S., & Dutta, K. 2025, DESI results: Hint towards coupled dark matter and dark energy, arXiv e-prints, arXiv:2503.10806, doi: 10.48550/arXiv.2503.10806

[48] Chaussidon, E., Yèche, C., de Mattia, A., et al. 2024, Constraining primordial non-Gaussianity with DESI 2024 LRG and QSO samples, arXiv e-prints, arXiv:2411.17623, doi: 10.48550/arXiv.2411.17623

[49] Chen, R., Scolnic, D., Rozo, E., et al. 2022, Measuring Cosmological Parameters with Type Ia Supernovae in redMaGiC Galaxies, The Astrophysical Journal, 938, 62, doi: 10.3847/1538-4357/ac8b82

[50] Chen, R., Scolnic, D., Vincenzi, M., et al. 2024, Evaluating Cosmological Biases Using Photometric Redshifts for Type Ia Supernova Cosmology with the Dark Energy Survey Supernova Program, doi: 10.48550/arXiv.2407.16744

[51] Chevallier, M., & Polarski, D. 2001, Accelerating Universes with Scaling Dark Matter, International Journal of Modern Physics D, 10, 213, doi: 10.1142/S0218271801000822

[52] Chisari, N. E., Alonso, D., Krause, E., et al. 2019, Core Cosmology Library: Precision Cosmological Predictions for LSST, ApJS, 242, 2, doi: 10.3847/1538-4365/ab1658

[53] Chudaykin, A., & Kunz, M. 2024, Modified gravity interpretation of the evolving dark energy in light of DESI data, Phys. Rev. D, 110, 123524, doi: 10.1103/PhysRevD.110.123524

[54] Ćiprijanović, A., Kafkes, D., Downey, K., et al. 2021, DeepMerge - II. Building Robust Deep Learning Algorithms for Merging Galaxy Identification across Domains, Monthly Notices of the Royal Astronomical Society, 506, 677, doi: 10.1093/mnras/stab1677

[55] Ćiprijanović, A., Kafkes, D., Snyder, G., et al. 2022, DeepAdversaries: Examining the Robustness of Deep Learning Models for Galaxy Morphology Classification, Machine Learning: Science and Technology, 3, 035007, doi: 10.1088/2632-2153/ac7f1a

[56] Cirelli, M., Strumia, A., & Zupan, J. 2024, Dark Matter, arXiv e-prints, arXiv:2406.01705, doi: 10.48550/arXiv.2406.01705

[57] Claver, C. F., & LSST Systems Engineering Integrated Project Team. 2022, Vera C. Rubin Observatory LSST System Requirements, http://ls.st/LSE-29

[58] Coe, D. 2009, Fisher Matrices and Confidence Ellipses: A Quick-Start Guide and Software, doi: 10.48550/arXiv.0906.4123

[59] Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology, The Astronomical Journal, 132, 926, doi: 10.1086/505530

[60] Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, Colors and magnitudes predicted for high redshift galaxies, The Astrophysical Journal Supplement Series, 43, 393, doi: 10.1086/190674

[61] Collaboration, DESI., Adame, A. G., Aguilar, J., et al. 2025, DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman Alpha Forest, J. Cosmol. Astropart. Phys., 2025, 124, doi: 10.1088/1475-7516/2025/01/124

[62] Collaboration, DESI., Lodha, K., Calderon, R., et al. 2025, Extended Dark Energy Analysis Using DESI DR2 BAO Measurements, arXiv, doi: 10.48550/arXiv.2503.14743

[63] Connolly, A. J., Csabai, I., Szalay, A. S., et al. 1995, Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry, The Astronomical Journal, 110, 2655, doi: 10.1086/117720

[64] Cooley, J., Lin, T., Lippincott, W. H., et al. 2022, Report of the Topical Group on Particle Dark Matter for Snowmass 2021, arXiv e-prints, arXiv:2209.07426, doi: 10.48550/arXiv.2209.07426

[65] Cooper, M. C., Aird, J. A., Coil, A. L., et al. 2011, The DEEP3 Galaxy Redshift Survey: Keck/DEIMOS spectroscopy in the goods-n field, Astrophysical Journal, Supplement Series, 193, 14, doi: 10.1088/0067-0049/193/1/14

[66] Cowperthwaite, P. S., Villar, V. A., Scolnic, D. M., & Berger, E. 2019, LSST Target-of-opportunity Observations of Gravitational-wave Events: Essential and Efficient, ApJ, 874, 88, doi: 10.3847/1538-4357/ab07b6

[67] Craig, N., Green, D., Meyers, J., & Rajendran, S. 2024, No νs is Good News, Journal of High Energy Physics, 2024, 97, doi: 10.1007/JHEP09(2024)097

[68] Crenshaw, J. F. 2024, Notes on Wavefront Estimation, Vera C. Rubin Observatory. https://sitcomtn-111.lsst.io

[69] Crenshaw, J. F., & Connolly, A. J. 2020, Learning Spectral Templates for Photometric Redshift Estimation from Broadband Photometry, The Astronomical Journal, 160, 191, doi: 10.3847/1538-3881/abb0e2

[70] Crenshaw, J. F., Kalmbach, J. B., Gagliano, A., et al. 2024, Probabilistic Forward Modeling of Galaxy Catalogs with Normalizing Flows, The Astronomical Journal, 168, 80, doi: 10.3847/1538-3881/ad54bf

[71] Crenshaw, J. F., Connolly, A. J., Meyers, J. E., et al. 2024, Using AI for Wave-front Estimation with the Rubin Observatory Active Optics System, The Astronomical Journal, 167, 86, doi: 10.3847/1538-3881/ad1661

[72] Crenshaw, J. F., Leistedt, B., Graham, M. L., et al. 2025, Quantifying the Impact of LSST u-band Survey Strategy on Photometric Redshift Estimation and the Detection of Lyman-break Galaxies, arXiv e-prints, arXiv:2503.06016, doi: 10.48550/arXiv.2503.06016

[73] Cross, D. N., & Sánchez, C. 2024, Inverse Galaxy-Galaxy Lensing: Magnification, Intrinsic Alignments and Cosmology, doi: 10.48550/arXiv.2410.00714

[74] Csabai, I., Connolly, A. J., Szalay, A. S., & Budavári, T. 2000, Reconstructing Galaxy Spectral Energy Distributions from Broadband Photometry,, Tech. Rep. 1, doi: 10.1086/301159

[75] Danielsson, U. H., & Riet, T. V. 2018, What If String Theory Has No de Sitter Vacua? International Journal of Modern Physics D, 27, 1830007, doi: 10.1142/S0218271818300070

[76] Das, S., Corasaniti, P. S., & Khoury, J. 2006, Superacceleration as the signature of a dark sector interaction, Phys. Rev. D, 73, 083509, doi: 10.1103/PhysRevD.73.083509

[77] Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, the Sdss-Iv Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data, The Astronomical Journal, 151, 44, doi: 10.3847/0004-6256/151/2/44

[78] de Jong, J. T., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, The Kilo-Degree Survey, Experimental Astronomy, 35, 25, doi: 10.1007/s10686-012-9306-1

[79] DES Collaboration, Abbott, T. M. C., Acevedo, M., et al. 2024, The Dark Energy Survey: Cosmology Results with ∼1500 New High-redshift Type Ia Supernovae Using the Full 5 yr Data Set, ApJL, 973, L14, doi: 10.3847/2041-8213/ad6f9f

[80] DES Collaboration, Abbott, T. M. C., Aguena, M., et al. 2025, Dark Energy Survey Year 3 Results: Cosmological Constraints from Cluster Abundances, Weak Lensing, and Galaxy Clustering, arXiv e-prints, arXiv:2503.13632, doi: 10.48550/arXiv.2503.13632

[81] DES Collaboration, Abbott, T. M. C., Acevedo, M., et al. 2025, Dark Energy Survey: Implications for Cosmological Expansion Models from the Final DES Baryon Acoustic Oscillation and Supernova Data, arXiv, doi: 10.48550/arXiv.2503.06712

[82] DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design, https://arxiv.org/abs/1611.00036

[83] DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, The DESI Experiment Part II: Instrument Design, arXiv e-prints, arXiv:1611.00037, doi: 10.48550/arXiv.1611.00037

[84] DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2024, DESI 2024 V: Full-Shape Galaxy Clustering from Galaxies and Quasars, arXiv e-prints, arXiv:2411.12021, doi: 10.48550/arXiv.2411.12021

[85] DESI Collaboration, Karim, M. A., Aguilar, J., et al. 2025, DESI DR2 Results I: Baryon Acoustic Oscillations from the Lyman Alpha Forest, arXiv, doi: 10.48550/arXiv.2503.14739

[86] Dey, B., Newman, J. A., Andrews, B. H., et al. 2021, Re-calibrating Photometric Redshift Probability Distributions Using Feature-space Regression, arXiv e-prints, arXiv:2110.15209, doi: 10.48550/arXiv.2110.15209

[87] Dey, B., Zhao, D., Newman, J. A., et al. 2022, Calibrated Predictive Distributions via Diagnostics for Conditional Coverage, doi: 10.48550/arXiv.2205.14568

[88] Di Valentino, E., Linder, E. V., & Melchiorri, A. 2018, Vacuum phase transition solves the H_0 tension, PhRvD, 97, 043528, doi: 10.1103/PhysRevD.97.043528

[89] Di Valentino, E., Mena, O., Pan, S., et al. 2021, In the realm of the Hubble tension-a review of solutions, Classical and Quantum Gravity, 38, 153001, doi: 10.1088/1361-6382/ac086d

[90] Dinh, L., Krueger, D., & Bengio, Y. 2015, in Proceedings of the 3rd International Conference on Learning Representations, ed. Y. Bengio & Y. LeCun, San Diego, CA. https://arxiv.org/abs/1410.8516

[91] Dinh, L., Sohl-Dickstein, J., & Bengio, S. 2017, in Proceedings of the 5th International Conference on Learning Representations, Toulon, France. https://arxiv.org/abs/1605.08803

[92] Donnan, C. T., McLeod, D. J., McLure, R. J., et al. 2023, The Abundance of $z \gtrsim 10$ Galaxy Candidates in the HUDF Using Deep JWST NIRCam Medium-Band Imaging, Monthly Notices of the Royal Astronomical Society, 520, 4554, doi: 10.1093/mnras/stad471

[93] Donnan, C. T., McLeod, D. J., Dunlop, J. S., et al. 2023, The Evolution of the Galaxy UV Luminosity Function at Redshifts z $\simeq$ 8 - 15 from Deep JWST and Ground-Based near-Infrared Imaging, Monthly Notices of the Royal Astronomical Society, 518, 6011, doi: 10.1093/mnras/stac3472

[94] Dunlop, J. S. 2013, in Astrophysics and Space Science Library, Vol. 396, The First Galaxies, ed. T. Wiklind, B. Mobasher, & V. Bromm, 223, doi: 10.1007/978-3-642-32362-1_5

[95] Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, in Advances in Neural Information Processing Systems 32, ed. H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, & R. Garnett (Vancouver, Canada: Curran Associates, Inc.), 7511–7522. https://arxiv.org/abs/1906.04032

[96] Ebina, H., & White, M. 2024, Cosmology before noon with multiple galaxy populations, JCAP, 2024, 052, doi: 10.1088/1475-7516/2024/06/052

[97] Efstathiou, G., Sutherland, W. J., & Maddox, S. J. 1990, The cosmological constant and cold dark matter, Nature, 348, 705, doi: 10.1038/348705a0

[98] Elbers, W., Aviles, A., Noriega, H. E., et al. 2025, Constraints on Neutrino Physics from DESI DR2 BAO and DR1 Full Shape, arXiv, doi: 10.48550/arXiv.2503.14744

[99] Ellerbroek, B., & Andersen, D. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7015, Adaptive Optics Systems, ed. N. Hubin, C. E. Max, & P. L. Wizinowich, 70155W, doi: 10.1117/12.788053

[100] Euclid Collaboration, Adam, R., Vannier, M., et al. 2019, Euclid preparation. III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection, A&A, 627, A23, doi: 10.1051/0004-6361/201935088

[101] Falorsi, L., de Haan, P., Davidson, T. R., & Forré, P. 2019, in Proceedings of Machine Learning Research, Vol. 89, The 22nd International Conference on Artificial Intelligence and Statistics, ed. K. Chaudhuri & M. Sugiyama (Naha, Okinawa, Japan: PMLR), 3244–3253. http://proceedings.mlr.press/v89/falorsi19a.html

[102] Farzan, Y., & Hannestad, S. 2016, Neutrinos secretly converting to lighter particles to please both KATRIN and the cosmos, JCAP, 2016, 058, doi: 10.1088/1475-7516/2016/02/058

[103] Ferrari, A. G., Ballardini, M., Finelli, F., & Paoletti, D. 2025, Scalar-tensor gravity and DESI 2024 BAO data, PhRvD, 111, 083523, doi: 10.1103/PhysRevD.111.083523

[104] Finkelstein, S. L., & Bagley, M. B. 2022, On the Coevolution of the AGN and Star-forming Galaxy Ultraviolet Luminosity Functions at $3 < z < 9$, The Astrophysical Journal, 938, 25, doi: 10.3847/1538-4357/ac89eb

[105] Finkelstein, S. L., Ryan, Russell E., J., Papovich, C., et al. 2015, The Evolution of the Galaxy Rest-frame Ultraviolet Luminosity Function over the First Two Billion Years, ApJ, 810, 71, doi: 10.1088/0004-637X/810/1/71

[106] Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2022, CEERS Key Paper I: An Early Look into the First 500 Myr of Galaxy Formation with JWST, arXiv, doi: 10.48550/arXiv.2211.05792

[107] Fort, S., Hu, H., & Lakshminarayanan, B. 2020, Deep Ensembles: A Loss Landscape Perspective, arXiv:1912.02757 [cs, stat]. https://arxiv.org/abs/1912.02757

[108] Fried, D. L. 1965, Statistics of a Geometric Representation of Wavefront Distortion, Journal of the Optical Society of America (1917-1983), 55, 1427. https://ui.adsabs.harvard.edu/abs/1965JOSA...55.1427F

[109] Fruchter, A., & Hook, R. 2002, Drizzle: A Method for the Linear Reconstruction of Undersampled Images,, Tech. Rep. 792, doi: 10.1086/338393

[110] Fukuda, Y., Hayakawa, T., Ichihara, E., et al. 1998, Evidence for Oscillation of Atmospheric Neutrinos, PhRvL, 81, 1562, doi: 10.1103/PhysRevLett.81.1562

[111] Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, The Gaia Mission, Astronomy and Astrophysics, 595, A1, doi: 10.1051/0004-6361/201629272

[112] Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, Gaia Data Release 2. Summary of the Contents and Survey Properties, Astronomy and Astrophysics, 616, A1, doi: 10.1051/0004-6361/201833051

[113] Gemici, M. C., Rezende, D. J., & Mohamed, S. 2016, Normalizing Flows on Riemannian Manifolds, CoRR, abs/1611.02304. https://arxiv.org/abs/1611.02304

[114] Giarè, W., Sabogal, M. A., Nunes, R. C., & Di Valentino, E. 2024, Interacting Dark Energy after DESI Baryon Acoustic Oscillation Measurements, Phys. Rev. Lett., 133, 251003, doi: 10.1103/PhysRevLett.133.251003

[115] Giavalisco, M. 2002, Lyman-Break Galaxies, ARA&A, 40, 579, doi: 10.1146/annurev.astro.40.121301.111837

[116] Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., et al. 2014, How Sensitive Are Predicted Galaxy Luminosities to the Choice of Stellar Population Synthesis Model? Monthly Notices of the Royal Astronomical Society, 439, 264, doi: 10.1093/mnras/stt2410

[117] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., et al. 2014, Generative Adversarial Networks, CoRR, abs/1406.2661, doi: 10.48550/arXiv.1406.2661

[118] Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, ApJ, 622, 759, doi: 10.1086/427976

[119] Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, Photometric Redshifts with the LSST: Evaluating Survey Observing Strategies, The Astronomical Journal, 155, 1, doi: 10.3847/1538-3881/aa99d4

[120] Graham, M. L., Connolly, A. J., Wang, W., et al. 2020, Photometric Redshifts with the LSST. II. The Impact of Near-infrared and Near-ultraviolet Photometry, The Astronomical Journal, 159, 258, doi: 10.3847/1538-3881/ab8a43

[121] Grazian, A., Giallongo, E., Boutsia, K., et al. 2024, What Are the Pillars of Reionization? Revising the AGN Luminosity Function at z $\sim$ 5, ApJ, 974, 84, doi: 10.3847/1538-4357/ad6980

[122] Green, D., Kaplan, D. E., & Rajendran, S. 2021, Neutrino interactions in the late universe, Journal of High Energy Physics, 2021, doi: 10.1007/jhep11(2021)162

[123] Green, J., Schechter, P., Baltay, C., et al. 2012, Wide-Field InfraRed Survey Telescope (WFIRST) Final Report, arXiv e-prints. https://arxiv.org/abs/1208.4012

[124] Groves, B., Brinchmann, J., & Walcher, C. J. 2012, The Balmer decrement of Sloan Digital Sky Survey galaxies, Monthly Notices of the Royal Astronomical Society, 419, 1402, doi: 10.1111/j.1365-2966.2011.19796.x

[125] Hamana, T., Miyazaki, S., Okura, Y., Okamura, T., & Futamase, T. 2013, Toward Understanding the Anisotropic Point Spread Function of Suprime-Cam and Its Impact on Cosmic Shear Measurement, Publications of the Astronomical Society of Japan, 65, 104, doi: 10.1093/pasj/65.5.104

[126] Hambleton, K. M., Bianco, F. B., Street, R., et al. 2023, Rubin Observatory LSST Transients and Variable Stars Roadmap, Publications of the Astronomical Society of the Pacific, 135, 105002, doi: 10.1088/1538-3873/acdb9a

[127] Hang, Q., Joachimi, B., Charles, E., et al. 2024, Impact of Survey Spatial Variability on Galaxy Redshift Distributions and the Cosmological 3 × 2-Point Statistics for the Rubin Legacy Survey of Space and Time (LSST), Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/stae2519

[128] Harikane, Y., Ouchi, M., Ono, Y., et al. 2018, GOLDRUSH. II. Clustering of galaxies at z $\sim$ 4-6 revealed with the half-million dropouts over the 100 deg^2 area corresponding to 1 Gpc^3, PASJ, 70, S11, doi: 10.1093/pasj/psx097

[129] Harikane, Y., Ono, Y., Ouchi, M., et al. 2022, GOLDRUSH. IV. Luminosity Functions and Clustering Revealed with 4,000,000 Galaxies at z 2-7: Galaxy-AGN Transition, Star Formation Efficiency, and Implication for Evolution at z $>$ 10, The Astrophysical Journal Supplement Series, 259, 20, doi: 10.3847/1538-4365/ac3dfc

[130] Harikane, Y., Ouchi, M., Oguri, M., et al. 2023, A Comprehensive Study of Galaxies at z 9-16 Found in the Early JWST Data: Ultraviolet Luminosity Functions and Cosmic Star Formation History at the Pre-reionization Epoch, The Astrophysical Journal Supplement Series, 265, 5, doi: 10.3847/1538-4365/acaaa9

[131] He, K., Zhang, X., Ren, S., & Sun, J. 2016, in CVPR 2016 (Las Vegas, NV, USA: IEEE Computer Society), 770–778, doi: 10.1109/CVPR.2016.90

[132] Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, PHAT: PHoto-z Accuracy Testing, A&A, 523, A31, doi: 10.1051/0004-6361/201014885

[133] Hill, J. C., McDonough, E., Toomey, M. W., & Alexander, S. 2020, Early dark energy does not restore cosmological concordance, PhRvD, 102, 043507, doi: 10.1103/PhysRevD.102.043507

[134] Ho, J., Chen, X., Srinivas, A., Duan, Y., & Abbeel, P. 2019, in Proceedings of Machine Learning Research, Vol. 97, Proceedings of the 36th International Conference on Machine Learning, ed. K. Chaudhuri & R. Salakhutdinov (Long Beach, CA, USA: PMLR), 2722–2730. http://proceedings.mlr.press/v97/ho19a.html

[135] Homar, G. M. 2023, The Rubin Observatory Optical Feedback Controller,, Tech. rep.

[136] Hoogeboom, E., Cohen, T. S., & Tomczak, J. M. 2020, Learning Discrete Distributions by Dequantization, CoRR, abs/2001.11235. https://arxiv.org/abs/2001.11235

[137] Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, Dark Energy Survey Year 1 Results: Redshift Distributions of the Weak-Lensing Source Galaxies, Monthly Notices of the Royal Astronomical Society, 478, 592, doi: 10.1093/mnras/sty957

[138] Hu, W. 2005, Crossing the phantom divide: Dark energy internal degrees of freedom, PhRvD, 71, 047301, doi: 10.1103/PhysRevD.71.047301

[139] Hudelot, P., Cuillandre, J.-C., Withington, K., et al. 2012, VizieR Online Data Catalog: The CFHTLS Survey (T0007 release) (Hudelot+ 2012), VizieR Online Data Catalog, II/317. http://adsabs.harvard.edu/abs/2012yCat.2317....0H

[140] Ilbert, O., Capak, P., Salvato, M., et al. 2009, Cosmos photometric redshifts with 30-bands for 2-deg2, Astrophysical Journal, 690, 1236, doi: 10.1088/0004-637X/690/2/1236

[141] Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, An Updated Analytic Model

for Attenuation by the Intergalactic Medium, Monthly Notices of the Royal Astronomical Society, 442, 1805, doi: 10.1093/mnras/stu936

[142] Ishikawa, S., Kashikawa, N., Toshikawa, J., et al. 2017, The Galaxy-Halo Connection in High-redshift Universe: Details and Evolution of Stellar-to-halo Mass Ratios of Lyman Break Galaxies on CFHTLS Deep Fields, ApJ, 841, 8, doi: 10.3847/1538-4357/aa6d64

[143] Ivanov, M. M., McDonough, E., Hill, J. C., et al. 2020, Constraining early dark energy with large-scale structure, PhRvD, 102, 103502, doi: 10.1103/PhysRevD.102.103502

[144] Ivezić, v., & the LSST Science Collaboration. 2018, The LSST System Science Requirements Document, Vera C. Rubin Observatory. https://docushare.lsst.org/docushare/dsweb/Get/LPM-17

[145] Ivezić, Ž., & LSST Science Collaboration. 2018, The LSST Science Requirements Document, LSST Project Management LPM-17. http://ls.st/srd

[146] Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, LSST: From Science Drivers to Reference Design and Anticipated Data Products, The Astrophysical Journal, 873, 111, doi: 10.3847/1538-4357/ab042c

[147] Iye, M., Noguchi, T., Torii, Y., Mikami, Y., & Ando, H. 1991, Evaluation of Seeing on a 62-Cm Mirror, Publications of the Astronomical Society of the Pacific, 103, 712, doi: 10.1086/132870

[148] Izbicki, R., & Lee, A. B. 2017, Converting high-dimensional regression to high-dimensional conditional density estimation, Electronic Journal of Statistics, 11, 2800, doi: 10.1214/17-EJS1302

[149] Izotov, Y. I., Guseva, N. G., Fricke, K. J., et al. 2021, Low-redshift compact star-forming galaxies as analogues of high-redshift star-forming galaxies, A&A, 646, A138, doi: 10.1051/0004-6361/202039772

[150] Jaeckel, J., Rybka, G., & Winslow, L. 2022, Report of the Topical Group on Wave Dark Matter for Snowmass 2021, arXiv e-prints, arXiv:2209.08125, doi: 10.48550/arXiv.2209.08125

[151] Jaini, P., Kobyzev, I., Yu, Y., & Brubaker, M. 2020, in Proceedings of Machine Learning Research, Vol. 119, arXiv:1907.04481 [Cs, Math, Stat] (Virtual: PMLR), 4673–4681. https://arxiv.org/abs/1907.04481

[152] Janish, R. J. 2012, PhD thesis, MIT, Cambridge, MA. http://hdl.handle.net/1721.1/78543

[153] Jarvis, M., Schechter, P., & Jain, B. 2008, Telescope Optics and Weak Lensing: PSF Patterns Due to Low Order Aberrations, arXiv e-prints, arXiv:0810.0027, doi: 10.48550/arXiv.0810.0027

[154] Jones, R. L., Yoachim, P., Ivezić, v., Neilsen, E. H. J., & Ribeiro, T. 2021, Survey Strategy and Cadence Choices for the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST), Vera C. Rubin Observatory. https://pstn-051.lsst.io

[155] Jones, R. L., Yoachim, P., Chandrasekharan, S., et al. 2014, in Observatory Operations: Strategies, Processes, and Systems V, ed. A. B. Peck, C. R. Benn, & R. L. Seaman, Vol. 9149, International Society for Optics and Photonics (SPIE), 91490B, doi: 10.1117/12.2056835

[156] Kalmbach, J. B., VanderPlas, J. T., & Connolly, A. J. 2020, Applying Information Theory to Design Optimal Filters for Photometric Redshifts, The Astrophysical Journal, 890, 74, doi: 10.3847/1538-4357/ab684f

[157] Katrin Collaboration, Aker, M., Batzler, D., et al. 2025, Direct neutrino-mass measurement based on 259 days of KATRIN data, Science, 388, 180, doi: 10.1126/science.adq9592

[158] Kilbinger, M. 2015, Cosmology with Cosmic Shear Observations: A Review, Reports on Progress in Physics, 78, 086901, doi: 10.1088/0034-4885/78/8/086901

[159] Kind, M. C., & Brunner, R. J. 2013, TPZ: Photometric redshift PDFs and ancillary information by using prediction trees and random forests, Monthly Notices of the Royal Astronomical Society, 432, 1483, doi: 10.1093/mnras/stt574

[160] Kingma, D. P., & Ba, J. 2015, in 3rd International Conference on Learning Representations, ed. Y. Bengio & Y. LeCun, San Diego, CA. http://arxiv.org/abs/1412.6980

[161] Kingma, D. P., & Welling, M. 2014, in 2nd International Conference on Learning Representations, ed. Y. Bengio & Y. LeCun, Banff, AB, Canada, doi: 10.48550/arXiv.1312.6114

[162] Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and Their Application to K-Corrections, The Astrophysical Journal, 467, 38, doi: 10.1086/177583

[163] Kobyzev, I., Prince, S. J. D., & Brubaker, M. A. 2020, Normalizing Flows: An Introduction and Review of Current Methods, IEEE Trans. Pattern Anal. Mach. Intell., 1, doi: 10.1109/TPAMI.2020.2992934

[164] Kofman, L. A., Gnedin, N. Y., & Bahcall, N. A. 1993, Cosmological Constant, COBE Cosmic Microwave Background Anisotropy, and Large-Scale Clustering, ApJ, 413, 1, doi: 10.1086/172970

[165] Korytov, D., Hearin, A., Kovacs, E., et al. 2019, CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST, The Astrophysical Journal Supplement Series, 245, 26, doi: 10.3847/1538-4365/ab510c

[166] Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, The Fourth Data Release of the Kilo-Degree Survey: Ugri Imaging and Nine-Band Optical-IR Photometry over 1000 Square Degrees, A&A, 625, A2, doi: 10.1051/0004-6361/201834918

[167] Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019, Large Ly α Opacity Fluctuations and Low CMB τ in Models of Late Reionization with Large Islands of Neutral Hydrogen Extending to z $<$ 5.5, Monthly Notices of the Royal Astronomical Society, 485, L24, doi: 10.1093/mnrasl/slz025

[168] Lakshminarayanan, B., Pritzel, A., & Blundell, C. 2017, in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, ed. I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett, 6402–6413. https://proceedings.neurips.cc/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html

[169] Lamman, C., Tsaprazi, E., Shi, J., et al. 2024, The IA Guide: A Breakdown of Intrinsic Alignment Formalisms, The Open Journal of Astrophysics, 7, 14, doi: 10.21105/astro.2309.08605

[170] Le Fèvre, O., Mellier, Y., McCracken, H. J., et al. 2004, The VIRMOS deep imaging survey, i. Overview, survey strategy, and cfh12k observations, Astronomy and Astrophysics, 417, 839, doi: 10.1051/0004-6361:20031767

[171] Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, The VIMOS VLT deep survey final data release: A spectroscopic sample of 35 016 galaxies and AGN out to z 6.7 selected with 17.5 iAB 24.75, Astronomy and Astrophysics, 559, 14, doi: 10.1051/0004-6361/201322179

[172] Lee, M. A., Budavári, T., Sullivan, I. S., & Connolly, A. J. 2019, Sub-Band Image Reconstruction Using Differential Chromatic Refraction, The Astronomical Journal, 157, 182, doi: 10.3847/1538-3881/ab139f

[173] Lemos, P., Coogan, A., Hezaveh, Y., & Perreault-Levasseur, L. 2023, Sampling-Based

Accuracy Testing of Posterior Estimators for General Inference, 40th International Conference on Machine Learning, 202, 19256, doi: 10.48550/arXiv.2302.03026

[174] Leonard, C. D., Rau, M. M., & Mandelbaum, R. 2024, Photometric Redshifts and Intrinsic Alignments: Degeneracies and Biases in 3$\times$2pt Analysis, arXiv. https://arxiv.org/abs/2401.06060

[175] Levi, M., Allen, L. E., Raichoor, A., et al. 2019, in Bulletin of the American Astronomical Society, Vol. 51, 57, doi: 10.48550/arXiv.1907.10688

[176] Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, The zCosmos 10k-bright spectroscopic sample, Astrophysical Journal, Supplement Series, 184, 218, doi: 10.1088/0067-0049/184/2/218

[177] Linder, E. V. 2003, Exploring the Expansion History of the Universe, PhRvL, 90, 091301, doi: 10.1103/PhysRevLett.90.091301

[178] Linder, E. V. 2021, The Rise of Dark Energy, arXiv e-prints, arXiv:2106.09581, doi: 10.48550/arXiv.2106.09581

[179] Linder, E. V. 2024, Interpreting Dark Energy Data Away from Λ, arXiv e-prints, arXiv:2410.10981, doi: 10.48550/arXiv.2410.10981

[180] Liu, F. T., Ting, K. M., & Zhou, Z.-H. 2012, Isolation-Based Anomaly Detection, ACM Trans. Knowl. Discov. Data, 6, 39, doi: 10.1145/2133360.2133363

[181] Lochner, M., Scolnic, D. M., Awan, H., et al. 2018, Optimizing the LSST Observing Strategy for Dark Energy Science: DESC Recommendations for the Wide-Fast-Deep Survey, arXiv e-prints, arXiv:1812.00515, doi: 10.48550/arXiv.1812.00515

[182] Lochner, M., Scolnic, D., Almoubayyed, H., et al. 2022, The Impact of Observing Strategy on Cosmological Constraints with LSST, ApJS, 259, 58, doi: 10.3847/1538-4365/ac5033

[183] Łokas, E. L., & Mamon, G. A. 2003, Dark matter distribution in the Coma cluster from galaxy kinematics: breaking the mass-anisotropy degeneracy, MNRAS, 343, 401, doi: 10.1046/j.1365-8711.2003.06684.x

[184] Lokken, M., Gagliano, A., Narayan, G., et al. 2022, The Simulated Catalogue of Optical Transients and Correlated Hosts (SCOTCH), https://ui.adsabs.harvard.edu/abs/2022arXiv220602815L

[185] Lorenz, C. S., Funcke, L., Calabrese, E., & Hannestad, S. 2019, Time-varying neutrino mass from a supercooled phase transition: Current cosmological constraints and impact on the Ω_m-σ_8 plane, PhRvD, 99, 023501, doi: 10.1103/PhysRevD.99.023501

[186] LoVerde, M., & Afshordi, N. 2008, Extended Limber approximation, PhRvD, 78, 123506, doi: 10.1103/PhysRevD.78.123506

[187] Loverde, M., & Weiner, Z. J. 2024, Massive Neutrinos and Cosmic Composition, Journal of Cosmology and Astroparticle Physics, 2024, 048, doi: 10.1088/1475-7516/2024/12/048

[188] LSST Dark Energy Science Collaboration, Abolfathi, B., Alonso, D., et al. 2021, The LSST DESC DC2 Simulated Sky Survey, The Astrophysical Journal Supplement Series, 253, 31, doi: 10.3847/1538-4365/abd62c

[189] LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, LSST Science Book, Version 2.0, arXiv e-prints, 0912, arXiv:0912.0201. http://adsabs.harvard.edu/abs/2009arXiv0912.0201L

[190] LSST Science Collaboration, Marshall, P., Anguita, T., et al. 2017, Science-Driven Optimization of the LSST Observing Strategy, arXiv e-prints, arXiv:1708.04058, doi: 10.48550/arXiv.1708.04058

[191] LSST Science Collaboration, Marshall, P., Anguita, T., et al. 2017, Science-Driven Optimization of the LSST Observing Strategy, arXiv e-prints, arXiv:1708.04058, doi: 10.48550/arXiv.1708.04058

[192] Ludwick, K. J. 2017, The viability of phantom dark energy: A review, Modern Physics Letters A, 32, 1730025, doi: 10.1142/S0217732317300257

[193] Mahajan, V. N. 1981, Zernike Annular Polynomials for Imaging Systems with Annular Pupils., Journal of the Optical Society of America (1917-1983), 71, 75. https://ui.adsabs.harvard.edu/abs/1981JOSA...71...75M

[194] Malkan, M. A., Cohen, D. P., Maruyama, M., et al. 2017, Lyman-Break Galaxies at z $\sim$ 3 in the Subaru Deep Field: Luminosity Function, Clustering, and [O III] Emission, The Astrophysical Journal, 850, 5, doi: 10.3847/1538-4357/aa9331

[195] Malz, A. I., Lanusse, F., Crenshaw, J. F., & Graham, M. L. 2021, An Information-Based Metric for Observing Strategy Optimization, Demonstrated in the Context of Photometric Redshifts with Applications to Cosmology, doi: 10.48550/arXiv.2104.08229

[196] Mandelbaum, R., Seljak, U., Hirata, C. M., et al. 2008, Precision Photometric Redshift Calibration for Galaxy-Galaxy Weak Lensing, Monthly Notices of the Royal Astronomical Society, 386, 781, doi: 10.1111/j.1365-2966.2008.12947.x

[197] Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, The Galaxy Evolution Explorer : A Space Ultraviolet Survey Mission ,, Tech. Rep. 1, doi: 10.1086/426387

[198] Mason, C. A., Trenti, M., & Treu, T. 2022, The Brightest Galaxies at Cosmic Dawn, https://ui.adsabs.harvard.edu/abs/2022arXiv220714808M

[199] Megias Homar, G., Kahn, S. M., Meyers, J. M., Crenshaw, J. F., & Thomas, S. J. 2024, The Active Optics System on the Vera C. Rubin Observatory: Optimal Control of Degeneracy among the Large Number of Degrees of Freedom, The Astrophysical Journal, 974, 108, doi: 10.3847/1538-4357/ad6cdc

[200] Menci, N., Adil, S. A., Mukhopadhyay, U., Sen, A. A., & Vagnozzi, S. 2024, Negative Cosmological Constant in the Dark Energy Sector: Tests from JWST Photometric and Spectroscopic Observations of High-Redshift Galaxies, Journal of Cosmology and Astroparticle Physics, 2024, 072, doi: 10.1088/1475-7516/2024/07/072

[201] Menci, N., Castellano, M., Santini, P., et al. 2022, High-Redshift Galaxies from Early JWST Observations: Constraints on Dark Energy Models, The Astrophysical Journal, 938, L5, doi: 10.3847/2041-8213/ac96e9

[202] Menci, N., Sen, A. A., & Castellano, M. 2024, The Excess of JWST Bright Galaxies: A Possible Origin in the Ground State of Dynamical Dark Energy in the Light of DESI 2024 Data, The Astrophysical Journal, 976, 227, doi: 10.3847/1538-4357/ad8d5b

[203] Merson, A. I., Baugh, C. M., Helly, J. C., et al. 2013, Lightcone Mock Catalogues from Semi-Analytic Models of Galaxy Formation - I. Construction and Application to the BzK Colour Selection, Monthly Notices of the Royal Astronomical Society, 429, 556, doi: 10.1093/mnras/sts355

[204] Meyers, J. E., Kirkby, D., & Thomas, D. 2019, Batoid,, [Computer Software] https://doi.org/10.11578/dc.20200708.1 doi: 10.11578/dc.20200708.1

[205] Mitra, A., Kessler, R., More, S., Hlozek, R., & LSST Dark Energy Science Collaboration. 2023, Using Host Galaxy Photometric Redshifts to Improve Cosmological Constraints with Type Ia Supernovae in the LSST Era, The Astrophysical Journal, 944, 212, doi: 10.3847/1538-4357/acb057

[206] Miyatake, H., Harikane, Y., Ouchi, M., et al. 2022, First Identification of a CMB Lensing Signal Produced by 1.5 Million Galaxies at z $\sim$4 : Constraints on Matter Density Fluctuations at High Redshift, Physical Review Letters, 129, 061301, doi: 10.1103/PhysRevLett.129.061301

[207] Miyazaki, S., Komiyama, Y., Sekiguchi, M., et al. 2002, Subaru prime focus camera - Suprime-Cam,, Tech. Rep. 6, doi: 10.1093/pasj/54.6.833

[208] Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, the 3D-HST Survey: Hubble Space Telescope WFC3/G141 Grism Spectra, Redshifts, and Emission Line Measurements for 100,000 Galaxies, The Astrophysical Journal Supplement Series, 225, 27, doi: 10.3847/0067-0049/225/2/27

[209] More, S., Sugiyama, S., Miyatake, H., et al. 2023, Hyper Suprime-Cam Year 3 results: Measurements of clustering of SDSS-BOSS galaxies, galaxy-galaxy lensing, and cosmic shear, PhRvD, 108, 123520, doi: 10.1103/PhysRevD.108.123520

[210] Moskowitz, I., Gawiser, E., Crenshaw, J. F., et al. 2024, Improving Photometric Redshift Estimates with Training Sample Augmentation, The Astrophysical Journal, 967, L6, doi: 10.3847/2041-8213/ad4039

[211] Moutard, T., Sawicki, M., Arnouts, S., et al. 2020, UV and U-band luminosity functions from CLAUDS and HSC-SSP - I. Using four million galaxies to simultaneously constrain the very faint and bright regimes to z $\sim$ 3, MNRAS, 494, 1894, doi: 10.1093/mnras/staa706

[212] Muir, J., Baxter, E., Miranda, V., et al. 2021, DES Y1 results: Splitting growth and geometry to test Λ CDM, PhRvD, 103, 023528, doi: 10.1103/PhysRevD.103.023528

[213] Munshi, D., Valageas, P., van Waerbeke, L., & Heavens, A. 2008, Cosmology with Weak Lensing Surveys, Physics Reports, 462, 67, doi: 10.1016/j.physrep.2008.02.003

[214] Naidu, R. P., Tacchella, S., Mason, C. A., et al. 2020, Rapid Reionization by the Oligarchs: The Case for Massive, UV-bright, Star-forming Galaxies with High Escape Fractions, The Astrophysical Journal, 892, 109, doi: 10.3847/1538-4357/ab7cc9

[215] Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022, Two Remarkably Luminous Galaxy Candidates at $z\approx11-13$ Revealed by JWST, doi: 10.48550/arXiv.2207.09434

[216] Nakane, M., Ouchi, M., Nakajima, K., et al. 2024, Lyα Emission at z = 7–13: Clear Evolution of Lyα Equivalent Width Indicating a Late Cosmic Reionization History, ApJ, 967, 28, doi: 10.3847/1538-4357/ad38c2

[217] Neill, D., Angeli, G., Claver, C., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9150, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 91500G, doi: 10.1117/12.2056553

[218] Newman, J. A., & Gruen, D. 2022, Photometric Redshifts for Next-Generation Surveys, Annu. Rev. Astron. Astrophys., 60, annurev, doi: 10.1146/annurev-astro-032122-014611

[219] Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts, Astrophysical Journal, Supplement Series, 208, 57, doi: 10.1088/0067-0049/208/1/5

[220] Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, Spectroscopic Needs for Imaging Dark Energy Experiments, Astroparticle Physics, 63, 81, doi: 10.1016/j.astropartphys.2014.06.007

[221] Niu, K., & Tian, C. 2022, Zernike Polynomials and Their Applications, Journal of Optics, 24, 123001, doi: 10.1088/2040-8986/ac9e08

[222] Noll, R. J. 1976, Zernike Polynomials and Atmospheric Turbulence*, J. Opt. Soc. Am., JOSA, 66, 207, doi: 10.1364/JOSA.66.000207

[223] Ono, Y., Ouchi, M., Harikane, Y., et al. 2018, Great Optically Luminous Dropout Research Using Subaru HSC (GOLDRUSH). I. UV Luminosity Functions at z $\sim$ 4-7 Derived

with the Half-Million Dropouts on the 100 Deg2 Sky, Publications of the Astronomical Society of Japan, 70, S10, doi: 10.1093/pasj/psx103

[224] Parker, L., & Raval, A. 2000, New quantum aspects of a vacuum-dominated universe, PhRvD, 62, 083503, doi: 10.1103/PhysRevD.62.083503

[225] Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, The galaxy UV luminosity function at z 2-4; new results on faint-end slope and the evolution of luminosity density, MNRAS, 456, 3194, doi: 10.1093/mnras/stv2857

[226] Payerne, C., d'Assignies Doumerg, W., Yèche, C., et al. 2024, High-Redshift LBG Selection from Broadband and Wide Photometric Surveys Using a Random Forest Algorithm, doi: 10.48550/arXiv.2410.08062

[227] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Scikit-Learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825. http://jmlr.org/papers/v12/pedregosa11a.html

[228] Peebles, P. J. E. 2020, Cosmology's Century: An Inside History of our Modern Understanding of the Universe, doi: 10.1515/9780691201665

[229] Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, Measurements of Ω and Λ from 42 High-Redshift Supernovae, ApJ, 517, 565, doi: 10.1086/307221

[230] Peterson, J. R., Jernigan, J. G., Kahn, S. M., et al. 2015, Simulation of Astronomical Images from Optical Survey Telescopes Using a Comprehensive Photon Monte Carlo Approach, The Astrophysical Journal Supplement Series, 218, 14, doi: 10.1088/0067-0049/218/1/14

[231] Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020, Planck 2018 results. IV. Diffuse component separation, A&A, 641, A4, doi: 10.1051/0004-6361/201833881

[232] Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, Planck 2018 Results. VI. Cosmological Parameters, Astronomy and Astrophysics, 641, A6, doi: 10.1051/0004-6361/201833910

[233] Planck Collaboration, Akrami, Y., Arroja, F., et al. 2020, Planck 2018 results. IX. Constraints on primordial non-Gaussianity, A&A, 641, A9, doi: 10.1051/0004-6361/201935891

[234] Poulin, V., Boddy, K. K., Bird, S., & Kamionkowski, M. 2018, Implications of an extended dark energy cosmology with massive neutrinos for cosmological tensions, PhRvD, 97, 123504, doi: 10.1103/PhysRevD.97.123504

[235] Poulin, V., Serpico, P. D., & Lesgourgues, J. 2016, A fresh look at linear cosmological constraints on a decaying Dark Matter component, JCAP, 2016, 036, doi: 10.1088/1475-7516/2016/08/036

[236] Poulin, V., Smith, T. L., Calderón, R., & Simon, T. 2025, Impact of ACT DR6 and DESI DR2 for Early Dark Energy and the Hubble tension, arXiv e-prints, arXiv:2505.08051, doi: 10.48550/arXiv.2505.08051

[237] Reddy, N. A., Steidel, C. C., Pettini, M., et al. 2008, Multiwavelength Constraints on the Cosmic Star Formation History from Spectroscopy: The Rest-Frame Ultraviolet, Hα, and Infrared Luminosity Functions at Redshifts 1.9 lesssim z lesssim 3.4, ApJS, 175, 48, doi: 10.1086/521105

[238] Rezende, D. J., Papamakarios, G., Racanière, S., et al. 2020, in Proceedings of Machine Learning Research, Vol. 119, arXiv:2002.02428 [Cs, Stat] (Virtual: PMLR), 8083–8092. https://arxiv.org/abs/2002.02428

[239] Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, AJ, 116, 1009, doi: 10.1086/300499

[240] Riess, A. G., Scolnic, D., Anand, G. S., et al. 2024, JWST Validates HST Distance Measurements: Selection of Supernova Subsample Explains Differences in JWST Estimates of Local H0, doi: 10.48550/arXiv.2408.11770

[241] Robertson, B. E., Banerji, M., Cooper, M. C., et al. 2017, Large Synoptic Survey Telescope Galaxies Science Roadmap, https://ui.adsabs.harvard.edu/abs/2017arXiv170801617R

[242] Roddier, C., & Roddier, F. 1993, Wave-Front Reconstruction from Defocused Images and the Testing of Ground-Based Optical Telescopes., Journal of the Optical Society of America A, 10, 2277, doi: 10.1364/JOSAA.10.002277

[243] Roddier, F. 1981, The Effects of Atmospheric Turbulence in Optical Astronomy, Progess in Optics, 19, 281, doi: 10.1016/S0079-6638(08)70204-X

[244] Roodman, A., Reil, K., & Davis, C. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9145 (Society of Photo-Optical Instrumentation Engineers (SPIE)), 914516, doi: 10.1117/12.2056904

[245] Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, GALSIM: The Modular Galaxy Image Simulation Toolkit, Astronomy and Computing, 10, 121, doi: 10.1016/j.ascom.2015.02.002

[246] Rubin, D., Aldering, G., Betoule, M., et al. 2023, Union Through UNITY: Cosmology with 2,000 SNe Using a Unified Bayesian Framework, arXiv e-prints, arXiv:2311.12098, doi: 10.48550/arXiv.2311.12098

[247] Rubin, V. C., & Ford, Jr., W. K. 1970, Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions, ApJ, 159, 379, doi: 10.1086/150317

[248] Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1978, Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa -¿ Sc., ApJL, 225, L107, doi: 10.1086/182804

[249] Rubin Observatory Project. 2025, An Interim Report on the ComCam On-Sky Campaign, Vera C. Rubin Observatory. https://sitcomtn-149.lsst.io

[250] Rubin Observatory Survey Cadence Optimization Committee. 2022, Survey Cadence Optimization Committee's Phase 1 Recommendations, Vera C. Rubin Observatory. https://pstn-053.lsst.io

[251] Rubin Observatory Survey Cadence Optimization Committee. 2023, Survey Cadence Optimization Committee's Phase 2 Recommendations, Vera C. Rubin Observatory. https://pstn-055.lsst.io

[252] Rubin Observatory Survey Cadence Optimization Committee. 2024, Survey Cadence Optimization Committee's Phase 3 Recommendations, Vera C. Rubin Observatory. https://pstn-056.lsst.io

[253] Ruhlmann-Kleider, V., Yèche, C., Magneville, C., et al. 2024, High Redshift LBGs from Deep Broadband Imaging for Future Spectroscopic Surveys, arXiv, doi: 10.48550/arXiv.2404.03569

[254] Sabti, N., Muñoz, J. B., & Blas, D. 2021, First Constraints on Small-Scale Non-Gaussianity from UV Galaxy Luminosity Functions, Journal of Cosmology and Astroparticle Physics, 2021, 010, doi: 10.1088/1475-7516/2021/01/010

[255] Sabti, N., Muñoz, J. B., & Blas, D. 2022, Galaxy Luminosity Function Pipeline for Cosmology and Astrophysics, Physical Review D, 105, 043518, doi: 10.1103/PhysRevD.105.043518

[256] Sabti, N., Muñoz, J. B., & Blas, D. 2022, New Roads to the Small-scale Universe: Measurements of the Clustering of Matter with the High-redshift UV Galaxy Luminosity Function, The Astrophysical Journal, 928, L20, doi: 10.3847/2041-8213/ac5e9c

[257] Sailer, N., Farren, G. S., Ferraro, S., & White, M. 2025, Dispuable: the high cost of a low optical depth, arXiv e-prints, arXiv:2504.16932, doi: 10.48550/arXiv.2504.16932

[258] Salcido, J., Bower, R. G., Barnes, L. A., et al. 2018, The Impact of Dark Energy on Galaxy Formation. What Does the Future of Our Universe Hold? Monthly Notices of the Royal Astronomical Society, 477, 3744, doi: 10.1093/mnras/sty879

[259] Salimans, T., Goodfellow, I. J., Zaremba, W., et al. 2016, in Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain, ed. D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, & R. Garnett, 2226–2234. https://proceedings.neurips.cc/paper/2016/hash/8a3363abe792db2d8761d6403605aeb7-Abstract.html

[260] Salvato, M., Ilbert, O., & Hoyle, B. 2019, The Many Flavours of Photometric Redshifts, Nature Astronomy, 3, 212, doi: 10.1038/s41550-018-0478-0

[261] Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, Photometric Redshift Analysis in the Dark Energy Survey Science Verification Data, Monthly Notices of the Royal Astronomical Society, 445, 1482, doi: 10.1093/mnras/stu1836

[262] Sawicki, M., Arnouts, S., Huang, J., et al. 2019, The CFHT Large Area U-band Deep Survey (CLAUDS), Monthly Notices of the Royal Astronomical Society, 489, 5202, doi: 10.1093/mnras/stz2522

[263] Scaramella, R., Amiaux, J., Mellier, Y., et al. 2022, Euclid Preparation - I. The Euclid Wide Survey, A&A, 662, A112, doi: 10.1051/0004-6361/202141938

[264] Schechter, P. L., & Sobel Levinson, R. 2011, Generic Misalignment Aberration Patterns in Wide-Field Telescopes, Publications of the Astronomical Society of the Pacific, 123, 812, doi: 10.1086/661111

[265] Schlegel, D. J., Ferraro, S., Aldering, G., et al. 2022, A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5, arXiv e-prints, arXiv:2209.03585, doi: 10.48550/arXiv.2209.03585

[266] Schlegel, D. J., Ferraro, S., Aldering, G., et al. 2022, A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5, arXiv e-prints, arXiv:2209.03585, doi: 10.48550/arXiv.2209.03585

[267] Schmidt, S. J., Newman, J. A., Abate, A., & Team, t. S. N. W. P. 2014, Spectroscopic Needs for Calibration of LSST Photometric Redshifts, https://arxiv.org/abs/1410.4506

[268] Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, Evaluation of Probabilistic Photometric Redshift Estimation Approaches for LSST, arXiv:2001.03621 [astro-ph]. https://arxiv.org/abs/2001.03621

[269] Schmittfull, M., & Seljak, U. 2018, Parameter Constraints from Cross-Correlation of CMB Lensing with Galaxy Clustering, Physical Review D, 97, 123540, doi: 10.1103/PhysRevD.97.123540

[270] Schneider, P. 2006, in , 269–451, doi: 10.1007/978-3-540-30310-7_3

[271] Schwamb, M. E., Jones, R. L., Chesley, S. R., et al. 2018, Large Synoptic Survey Telescope Solar System Science Roadmap, arXiv, doi: 10.48550/arXiv.1802.01783

[272] Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, The VIMOS public extragalactic redshift survey (VIPERS): Full spectroscopic data and auxiliary information release (PDR-2), Astronomy and Astrophysics, 609, doi: 10.1051/0004-6361/201630114

[273] Scott, B. R., Malz, A. I., & Sorba, R. 2024, A holistic exploration of the potentially recoverable redshift information of Stage IV galaxy surveys, arXiv e-prints, arXiv:2409.20443, doi: 10.48550/arXiv.2409.20443

[274] Sebag, J., & LSST Systems Engineering Integrated Project Team. 2016, Vera C. Rubin Observatory LSST Telescope Error Budget, http://ls.st/LTS-124

[275] Sen, M., & Smirnov, A. Y. 2024, Neutrinos with refractive masses and the DESI BAO results, arXiv e-prints, arXiv:2407.02462, doi: 10.48550/arXiv.2407.02462

[276] Seo, B.-J., Nissly, C., Angeli, G., et al. 2009, Analysis of Normalized Point Source Sensitivity as a Performance Metric for Large Telescopes, Appl. Opt., AO, 48, 5997, doi: 10.1364/AO.48.005997

[277] Sims, P. H., Bevins, H. T. J., Fialkov, A., et al. 2025, Rapid and Late Cosmic Reionization Driven by Massive Galaxies: a Joint Analysis of Constraints from 21-cm, Lyman Line & CMB Data Sets, arXiv e-prints, arXiv:2504.09725, doi: 10.48550/arXiv.2504.09725

[278] Sofue, Y., & Rubin, V. 2001, Rotation Curves of Spiral Galaxies, ARA&A, 39, 137, doi: 10.1146/annurev.astro.39.1.137

[279] Spergel, D., Gehrels, N., Baltay, C., et al. 2015, Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report, doi: 10.48550/arXiv.1503.03757

[280] Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Simulations of the formation, evolution and clustering of galaxies and quasars, Nature, 435, 629, doi: 10.1038/nature03597

[281] Stanway, E. R., Bremer, M. N., & Lehnert, M. D. 2008, On contamination and completeness in z ¿= 5 Lyman-break galaxy surveys, MNRAS, 385, 493, doi: 10.1111/j.1365-2966.2008.12853.x

[282] Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, Spectroscopic Confirmation of a Population of Normal Star-forming Galaxies at Redshifts Z > 3, The Astrophysical Journal, 462, L17, doi: 10.1086/310029

[283] Stephenson, P. C. L. 2014, Recurrence Relations for the Cartesian Derivatives of the Zernike Polynomials, J. Opt. Soc. Am. A, JOSAA, 31, 708, doi: 10.1364/JOSAA.31.000708

[284] Stevans, M. L., Finkelstein, S. L., Wold, I., et al. 2018, Bridging Star-forming Galaxy and AGN Ultraviolet Luminosity Functions at z = 4 with the SHELA Wide-field Survey, ApJ, 863, 63, doi: 10.3847/1538-4357/aacbd7

[285] Strauss, M. A., & Willick, J. A. 1995, The density and peculiar velocity fields of nearby galaxies, PhR, 261, 271, doi: 10.1016/0370-1573(95)00013-7

[286] Stylianou, N., Malz, A. I., Hatfield, P., Crenshaw, J. F., & Gschwend, J. 2022, The Sensitivity of GPz Estimates of Photo-z Posterior PDFs to Realistically Complex Training Set Imperfections, Publications of the Astronomical Society of the Pacific, 134, 044501, doi: 10.1088/1538-3873/ac59bf

[287] Tallis, M., Bailey, V. P., Macintosh, B., et al. 2020, Effects of mirror seeing on high-contrast adaptive optics instruments, Journal of Astronomical Telescopes, Instruments, and Systems, 6, 015002, doi: 10.1117/1.JATIS.6.1.015002

[288] Tang, X. T., Brout, D., Karwal, T., et al. 2025, Uniting the Observed Dynamical Dark Energy Preference with the Discrepancies in $Ω_m$ and H_0 Across Cosmological Probes, arXiv, doi: 10.48550/arXiv.2412.04430

[289] The Dark Energy Survey Collaboration. 2005, The Dark Energy Survey, arXiv e-prints, astro, doi: 10.48550/arXiv.astro-ph/0510346

[290] The Dark Energy Survey Collaboration. 2005, The Dark Energy Survey, arXiv e-prints, astro. https://arxiv.org/abs/0510346

[291] The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document, arXiv e-prints, 1809, arXiv:1809.01669. http://adsabs.harvard.edu/abs/2018arXiv180901669T

[292] Thomas, D., Meyers, J., & Kahn, S. M. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11448, Adaptive Optics Systems VII, ed. L. Schreiber, D. Schmidt, & E. Vernet, 114484H, doi: 10.1117/12.2576020

[293] Thomas, D., Meyers, J., & Kahn, S. M. 2021, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2076–2085

[294] Thomas, S., Tsai, T.-W., Sundararaman, H., et al. 2022, Vera C. Rubin Observatory Main Telescope Active Optics System: Wavefront Estimation Pipeline and Optical Feedback Controller Requirement Document, http://ls.st/LTS-186

[295] Thomas, S. J., Xin, B., Tsai, T.-W., et al. 2017, in AO4ELT5 Proceedings (Tenerife, Canary Islands, Spain: Instituto de Astrofísica de Canarias), doi: 10.26698/AO4ELT5.0137

[296] Thomas, S. J., Barr, J., Callahan, S., et al. 2022, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 12182, Ground-Based and Airborne Telescopes IX, ed. H. K. Marshall, J. Spyromilio, & T. Usuda, 121820W, doi: 10.1117/12.2630226

[297] Ting, K., Liu, F., & Zhou, Z. 2008, in ICDM 2008. Eighth IEEE International Conference on Data Mining (IEEE Computer Society), 413–422, doi: 10.1109/ICDM.2008.17

[298] Tokovinin, A. 2002, From Differential Image Motion to Seeing, Publications of the Astronomical Society of the Pacific, 114, 1156, doi: 10.1086/342683

[299] Tokovinin, A., & Heathcote, S. 2006, Donut: Measuring Optical Aberrations from a Single Extrafocal Image, Publications of the Astronomical Society of the Pacific, 118, 1165, doi: 10.1086/506972

[300] Toshikawa, J., Uchiyama, H., Kashikawa, N., et al. 2018, GOLDRUSH. III. A systematic search for protoclusters at z $\sim$ 4 based on the ¿100 deg^2 area, PASJ, 70, S12, doi: 10.1093/pasj/psx102

[301] Tudorica, A., Hildebrandt, H., Tewes, M., et al. 2017, Weak Lensing Magnification of SpARCS Galaxy Clusters, Astronomy and Astrophysics, 608, A141, doi: 10.1051/0004-6361/201731267

[302] Turner, M. S., KICP/UChicago, & Kavli Foundation, T. 2021, Understanding BBN: The Physics and Its History, arXiv e-prints, arXiv:2111.14254, doi: 10.48550/arXiv.2111.14254

[303] Valcin, D., Jimenez, R., Seljak, U., & Verde, L. 2025, The Age of the Universe with Globular Clusters III: Gaia Distances and Hierarchical Modeling, arXiv e-prints, arXiv:2503.19481, doi: 10.48550/arXiv.2503.19481

[304] van den Busch, J. L., Hildebrandt, H., Wright, A. H., et al. 2020, Testing KiDS Cross-Correlation Redshifts with Simulations, Astronomy and Astrophysics, 642, A200, doi: 10.1051/0004-6361/202038835

[305] Vogelsberger, M., Genel, S., Springel, V., et al. 2014, Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe, MNRAS, 444, 1518, doi: 10.1093/mnras/stu1536

[306] von Kármán, T. 1948, Progress in the Statistical Theory of Turbulence, Proceedings of the National Academy of Science, 34, 530, doi: 10.1073/pnas.34.11.530

[307] Vulcani, B., Trenti, M., Calvi, V., et al. 2017, Characterization and Modeling of Contamination for Lyman Break Galaxy Samples at High Redshift, The Astrophysical Journal, 836, 239, doi: 10.3847/1538-4357/aa5caf

[308] Weaverdyck, N., & Huterer, D. 2021, Mitigating Contamination in LSS Surveys: A Comparison of Methods, Monthly Notices of the Royal Astronomical Society, 503, 5061, doi: 10.1093/mnras/stab709

[309] Weinberg, S. 1987, Anthropic bound on the cosmological constant, PhRvL, 59, 2607, doi: 10.1103/PhysRevLett.59.2607

[310] Wes McKinney. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & Jarrod Millman, 56–61, doi: 10.25080/Majora-92bf1922-00a

[311] Wilson, A. G., & Izmailov, P. 2020, Bayesian Deep Learning and a Probabilistic Perspective of Generalization, arXiv:2002.08791 [cs, stat]. https://arxiv.org/abs/2002.08791

[312] Wilson, M. J., & White, M. 2019, Cosmology with Dropout Selection: Straw-Man Surveys &amp$\mathsemicolon$ CMB Lensing, J. Cosmol. Astropart. Phys., 2019, 015, doi: 10.1088/1475-7516/2019/10/015

[313] Winkler, C., Worrall, D., Hoogeboom, E., & Welling, M. 2019, Learning Likelihoods with Conditional Normalizing Flows, CoRR, abs/1912.00042. https://arxiv.org/abs/1912.00042

[314] Wolf, W. J., García-García, C., Bartlett, D. J., & Ferreira, P. G. 2024, Scant evidence for thawing quintessence, PhRvD, 110, 083528, doi: 10.1103/PhysRevD.110.083528

[315] Wright, A. H., Stölzner, B., Asgari, M., et al. 2025, KiDS-Legacy: Cosmological constraints from cosmic shear with the complete Kilo-Degree Survey, arXiv e-prints, arXiv:2503.19441, doi: 10.48550/arXiv.2503.19441

[316] Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2025, KiDS-Legacy: Redshift distributions and their calibration, arXiv e-prints, arXiv:2503.19440, doi: 10.48550/arXiv.2503.19440

[317] Xin, B., Claver, C., Liang, M., et al. 2015, Curvature Wavefront Sensing for the Large Synoptic Survey Telescope, Applied Optics, 54, 9045, doi: 10.1364/AO.54.009045

[318] Ye, G., Martinelli, M., Hu, B., & Silvestri, A. 2024, Non-minimally coupled gravity as a physically viable fit to DESI 2024 BAO, arXiv e-prints, arXiv:2407.15832, doi: 10.48550/arXiv.2407.15832

[319] Yin, J. E., Eisenstein, D. J., Finkbeiner, D. P., Stubbs, C. W., & Wang, Y. 2021, Active Optical Control with Machine Learning: A Proof of Concept for the Vera C. Rubin Observatory, AJ, 161, 216, doi: 10.3847/1538-3881/abe9b9

[320] Yu, B., Knight, R. Z., Sherwin, B. D., et al. 2018, Towards Neutrino Mass from Cosmology without Optical Depth Information, doi: 10.48550/arXiv.1809.02120

[321] Zhang, T., Rau, M. M., Mandelbaum, R., Li, X., & Moews, B. 2023, Photometric Redshift Uncertainties in Weak Gravitational Lensing Shear Analysis: Models and Marginalization, Monthly Notices of the Royal Astronomical Society, 518, 709, doi: 10.1093/mnras/stac3090

[322] Zhao, C., & Burge, J. H. 2007, Orthonormal Vector Polynomials in a Unit Circle, Part I: Basis Set Derived from Gradients of Zernike Polynomials, Opt. Express, OE, 15, 18014, doi: 10.1364/OE.15.018014

[323] Zhou, R., Cooper, M. C., Newman, J. A., et al. 2019, Deep ugrizY imaging and DEEP2/3 spectroscopy: a photometric redshift testbed for LSST and public release of data from the DEEP3 Galaxy Redshift Survey, Monthly Notices of the Royal Astronomical Society, 488, 4565, doi: 10.1093/mnras/stz1866

[324] Zuntz, J., Lanusse, F., Malz, A. I., et al. 2021, The LSST-DESC 3x2pt Tomography Optimization Challenge, The Open Journal of Astrophysics, 4, 13, doi: 10.21105/astro.2108.13418

[325] Zwicky, F. 1937, On the Masses of Nebulae and of Clusters of Nebulae, ApJ, 86, 217, doi: 10.1086/143864

Appendix A

THE RUBIN OBSERVATORY ACTIVE OPTICS SYSTEM

The Rubin Observatory is tasked with completing the Legacy Survey of Space and Time (LSST), the largest astronomical survey ever undertaken. To accomplish this goal, Rubin was designed as a three-mirror anastigmat with a fast optical system ($f/1.23$), engineered to deliver high optical quality over a huge $9.6\,\mathrm{deg}^2$ field of view. This image quality, however, is constantly imperiled by the changing gravitational load and evolving environmental conditions that perturb the optical system as the telescope slews across the sky, observing from sunset to sunrise, on nights throughout the year. These perturbations degrade the optical quality of the telescope, and, especially worrying for precision cosmology, these degradations are spatially correlated on the sky. Two primary drivers of optical aberrations are the direction of the gravity vector and the temperature of the system. The former is determined by the telescope elevation angle, which is correlated with declination, while the latter is correlated with the time of year, which is in turn correlated with right ascension. These aberrations modulate the delivered signal to noise, contaminating galaxy clustering signals [308], and bias galaxy shape measurements, contaminating cosmic shear [125, 153].

The Rubin Observatory employs an active optics system (AOS) to maintain optical alignment and mirror figure to reduce the impact of these perturbations and deliver the pristine optical quality required for precision science. Note that *active* optics, which correct optical alignment and mirror figure errors, operating at $\mathcal{O}(0.1\,\mathrm{Hz})$, should not be confused with *adaptive* optics, which correct for atmospheric seeing and operate at much higher frequencies. Rubin is not equipped with an *adaptive* optics system due to its wide field of view: light from sources across the field of view pass through different columns of air, making it

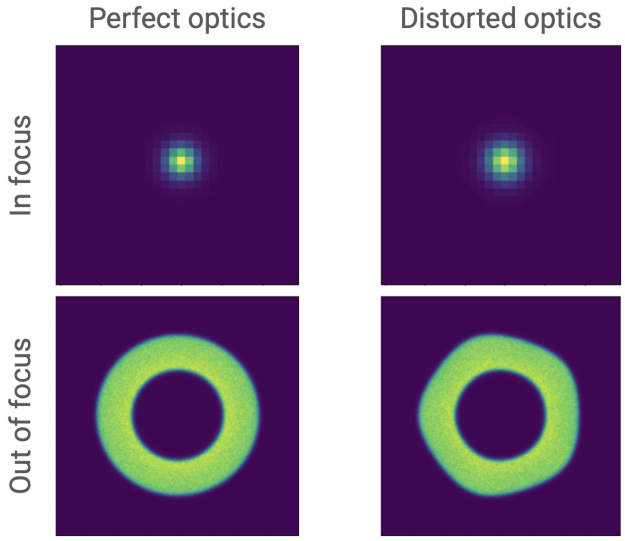

Figure A.1: Demonstrating the principle of out-of-focus wave-front sensing. In the in-focus images, you can see the right image has a larger PSF than the left image, however it is very difficult to diagnose the origin of this difference as the pixel size provides very little spatial resolution for the in-focus PSF. When we image out of focus, however, it is apparent that the right image has very large pentafoil aberration, associated with bending in the primary mirror.

impossible to apply a single correction to reduce seeing for all sources. [1]

Rubin's AOS consists of an open-loop and closed-loop component. The open-loop component uses a look-up table to provide predictable near-optimal corrections, using a model that is fit to the telescope elevation angle, the camera rotator angle, and the bulk mirror temperature. The closed-loop component corrects for the residual perturbations that result from temperature gradients within the system, hysteresis, wind loading, and other unpredictable environmental conditions. For the closed loop to operate it must be equipped with sensors that detect aberrations in the optical system.

[1] Active optics can, in principle, correct for ground-layer and dome-seeing turbulent modes that are stable on ~ 1 min time scales, if these exist.

There are curvature wavefront sensors at each corner of the Rubin focal plane to facilitate wavefront estimation for the Rubin closed loop. Each of these sensors is identical to a Rubin CCD sensor, except it is split in half, with half of the sensor 1.5 mm inside of focus, and the other half 1.5 mm outside of focus. Stars imaged on these sensors are out of focus, and therefore appear as images of the pupil, i.e. an annulus (typically referred to as "donuts" by the Rubin AOS team). These out-of-focus images provide the spatial resolution necessary to detail the imperfection in the telescope wavefront by spreading the PSF across a large number of pixels (Figure A.1). Wavefront estimation algorithms analyze these donut images to infer how the telescope wavefront departs from the ideal spherical form. This inference is then used to correct the optical alignment and the shape of the mirrors to improve the image quality in real-time [199]

The following appendices provide details of the wavefront estimation process. The rest of this appendix describes the physical components of the AOS (drawing from [14, 135, 199, 217, 295, 296, 317]). For more information about commissioning the system on the Rubin Commissioning Camera (ComCam), consult [249]. Information about commissioning the system on LSSTCam will be published in the near future.

A.1 The M1M3 mirror

Rubin has a three-mirror anastigmat design, including an 8.4 m primary mirror and 5.0 m tertiary mirror integrated into a single monolithic structure, composed of borosilicate glass. This unique design increases the compactness and stiffness of the system, and permanently fixes M3's position and orientation relative to M1, eliminating six degrees of freedom from the AOS, significantly simplifying optical alignment. The mirror sits on the M1M3 mirror cell, which comprises its support system, contains many sensors that provide telemetry about the state of the mirror, and houses 156 pneumatic actuators. These actuators provide active support to apply mirror figure corrections and to resist stressing the mirror with differential loads as the telescope slews to different positions on the sky.

The borosilicate glass of M1M3 has a relatively high coefficient of thermal expansion,

necessitating an extensive thermal control system. This system circulates air through the mirror's internal honeycomb structure, around its perimeter, and across the back surface to minimize thermal gradients within the mirror, which would otherwise stress the glass and significantly degrade image quality. The system also aims to keep the mirror slightly cooler than the air above it to reduce mirror seeing[2] [147, 287].

A.2 The M2 mirror

Rubin's 3.5 m secondary mirror is suspended above the M1M3, supported by the M2 mirror cell. The M2 is a thin meniscus of Ultra Low Expansion (ULE) glass, for which the coefficient of thermal expansion is only 1% that of borosilicate. As such the M2 mirror cell contains a comparatively minimal thermal control system, designed primarily to dissipate the heat generated by the active support system.

78 electromechanical actuators control the M2 mirror figure and provide active support to the glass: 72 in the axial direction and 6 in the tangential direction. The position and rotation of the M2 is controlled by a hexapod system, enabling the AOS to align the M2 with the M1M3.

A.3 LSSTCam

LSSTCam, Rubin's 3.2 gigapixel camera, is suspended on-axis below M2 and above M3. The camera is supported by a hexapod system that enables the AOS to align the camera with the mirrors. The camera also features a rotator to compensate for sky rotation during tracking of the sky. Due to slight tilts and misalignments in the camera lenses and focal plane, the camera is not perfectly rotationally symmetric, necessitating a camera rotator LUT in the AOS open loop.

[2]Mirror seeing is the generation of turbulence directly above a mirror due to thermal gradients between the mirror and the air.

A.4 50 Degrees of freedom

The above mechanical components are utilized by the AOS to maintain optical quality by controlling the following 50 degrees of freedom:

- 6 rigid-body motions of the M2 (decenter, piston, tip and tilt)
- 6 rigid-body motions of the camera (decenter, piston, tip and tilt)
- 20 bending modes of the M1M3
- 20 bending modes of the M2

The bending modes of the M1M3 and M2 were determined by finite-element analyses and the 20 most-important modes for controlling optical quality were selected for each.

The Rubin Observatory is the first wide-field telescope to control such a large number of degrees of freedom to maintain alignment and mirror figure. There are many degeneracies in such a complicated system. For example, in the presence of noise, the first two bending modes on M1M3 are highly degenerate with the first two bending modes on M2, as each cause astigmatism in the system. For a given optical aberration, therefore, the AOS must weigh many factors in determining which degree of freedom to use to correct for the aberration. In practice, rather than controlling degrees of freedom individually, we control linear combinations of them (so called "v modes"). Each mode is weighted by its contribution to the image quality, as well as the range of values given to the component degrees of freedom (e.g., the M2 is more flexible than the M1M3, so its bending modes have a larger allowed range of values). We also use a singular value decomposition (SVD) to truncate the most degenerate modes. See G. Megias Homar et al. [199] for more details.

Appendix B

NOTES ON WAVEFRONT ESTIMATION

This appendix was originally published by the Rubin Observatory as System Integration, Test, and Commissioning Technical Note 111 (SITCOMTN-111)[1], and provides an introduction to wavefront estimation, including the derivation of key equations used by the Rubin Observatory active optics wavefront estimation pipeline (WEP).

B.1 Introduction

The Rubin Observatory's Simonyi Survey Telescope must use an active optics system to correct perturbations to the optical alignment and mirror figures. Achieving these corrections requires the following steps:

1. estimating the wavefront of the optical system,

2. using the wavefront to estimate the current configuration of the optical system,

3. using the current optical configuration to derive forces to apply to the hexapods and actuators.

This note concerns the first step, filling in details behind B. Xin et al. [317] using the notation of version 9 of the wavefront estimation pipeline (WEP).

Estimating the wavefront typically requires an optical model for mapping the photons from the pupil to the image plane. Deriving this mapping is the focus of Section B.2. Section B.3 discusses different methods for estimating the wavefront of the telescope from defocused images.

[1] sitcomtn-111.lsst.io

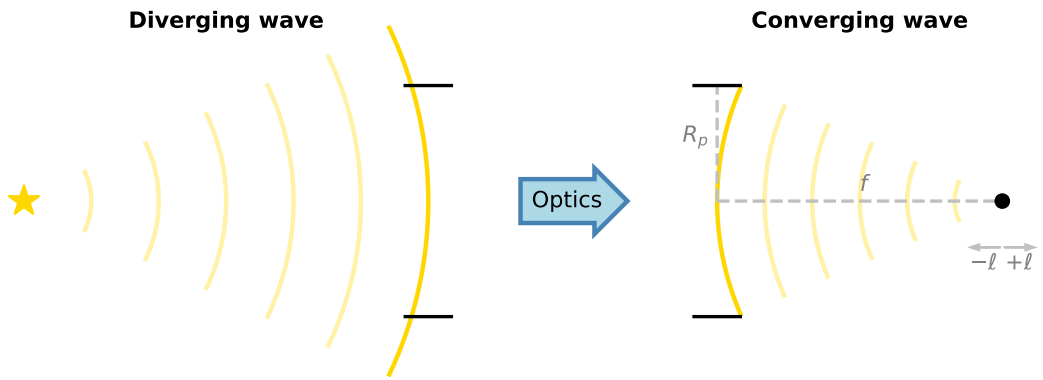

Figure B.1: Black box diagram of a telescope. An incoming diverging wave enters the pupil on the left, is transformed by the optical system, and is emitted from the pupil on the right, where it converges on the focal plane.

B.2 Mapping from the Pupil to the Image Plane

In this section we define the wavefront of the telescope and derive a mapping from photon positions on the pupil to the corresponding positions on the image plane.

B.2.1 The On-Axis and Paraxial Models

The optical system of a telescope can be treated as a black box which receives an incoming wave on the pupil plane and converts that wave to one converging on the focal plane (Fig. B.1). For example, ignoring phase perturbations from e.g. atmospheric turbulence, we can assume the incoming wave is a plane wave with phase

$$\varphi(z, \mathbf{x}) = -kz, \tag{B.1}$$

where k is the wavenumber and $\mathbf{x} = (x, y)$ is a 2D transverse position vector. After passing through the optical system, the wave emerges with the phase

$$\varphi(z, \mathbf{x}) = -k\sqrt{(z-f)^2 + \mathbf{x}^2}, \tag{B.2}$$

where f is the focal length of the system and we have defined $z = 0$ at the pupil.

We define the wavefront at the pupil using the reference sphere, which we define as the sphere with radius f that is centered on the focal plane and passes through the center of the pupil – i.e., the first emerging wavefront on the right of Fig. B.1. The points on the reference sphere are the set

$$\left\{(z_p, \mathbf{x}_p) \mid z_p = f - \sqrt{f^2 - \mathbf{x}_p^2}, \quad \mathbf{x}_p^2 < R_p^2\right\}, \tag{B.3}$$

where R_p is the radius of the pupil.

On the reference sphere the gradient of the phase is

$$\nabla\varphi(\mathbf{x}_p) = \frac{k}{f}\left(\sqrt{f^2 - \mathbf{x}_p^2}\,\hat{\mathbf{z}} - \mathbf{x}_p\right). \tag{B.4}$$

We have added the subscript p to the transverse vector x to remind that this equation applies to the wavefront at the pupil. Recall that photons propagate in the direction of the phase gradient. From Eq. B.4, you can see that a photon starting at $(z_p, \mathbf{x}_p)$ propagates in the direction $(f - z_p, -\mathbf{x}_p)$, and thus converges on the point $(f, 0)$. In other words, all photons emerging from the pupil converge on the focal point.

The above assumes a perfect optical system. We wish to consider an optical system with perturbations to alignment and mirror figure, which induce a phase perturbation $\delta\varphi = -k\,W(\mathbf{x}_p)$. $W(\mathbf{x}_p)$ is known as the optical path difference (OPD) and is defined on the reference sphere. With this perturbation, the phase gradient is now

$$\nabla\varphi(\mathbf{x}_p) = \frac{k}{f}\left(\sqrt{f^2 - \mathbf{x}_p^2}\,\hat{\mathbf{z}} - \mathbf{x}_p - f\nabla_{\mathbf{x}}W(\mathbf{x}_p)\right). \tag{B.5}$$

From Eq. B.5 you can see the effect of optical perturbations is to deflect photons from position $\mathbf{x}_p$ on the reference sphere by the amount $-f\nabla_{\mathbf{x}}W(\mathbf{x}_p)$ on the focal plane.

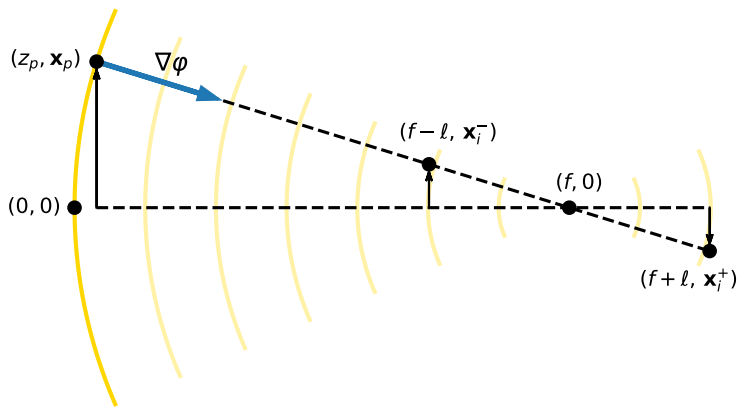

Figure B.2: Diagram showing the mapping of a pupil point to points on the defocused image planes. Note the line from $(z_p, \mathbf{x}_p)$ to $(f+\ell, \mathbf{x}_i^+)$ only passes through the focal point when $W(\mathbf{x}_p) = 0$. This geometry is the basis for Eq. B.6.

Active optics systems usually operate on defocused images. In particular, Rubin uses images from curvature wavefront sensors (CWFSs) that are offset on either side of the focal plane by a small distance ℓ. Thus we are interested in the images formed on the defocused planes at positions $z_i^\pm = f \pm \ell$. We can determine the mapping from pupil position $\mathbf{x}_p$ to image position $(\mathbf{x}_i^\pm, z_i^\pm)$ using the similar triangles in Fig. B.2:

$$\frac{\mathbf{x}_i^\pm - \mathbf{x}_p}{z_i^\pm - z_p} = \frac{\nabla_\mathbf{x}\varphi(\mathbf{x}_p)}{\partial\varphi(\mathbf{x}_p)/\partial z} \quad \Longrightarrow \quad \frac{\mathbf{x}_i^\pm - \mathbf{x}_p}{\sqrt{f^2 - \mathbf{x}_p} \pm \ell} = \frac{-\mathbf{x}_p - f\nabla_\mathbf{x} W(\mathbf{x}_p)}{\sqrt{f^2 - \mathbf{x}_p^2}}. \tag{B.6}$$

Solving for the image position yields

$$\mathbf{x}_i^\pm = \mp\frac{\ell}{\sqrt{f^2 - \mathbf{x}_p^2}}\,\mathbf{x}_p - f\left(1 \pm \frac{\ell}{\sqrt{f^2 - \mathbf{x}_p^2}}\right)\nabla_\mathbf{x} W(\mathbf{x}_p). \tag{B.7}$$

We see that for an unaberrated wavefront, the magnitude of the vector $\mathbf{x}_p$ shrinks linearly with respect to position on the z-axis. The effect of defocused-imaging is to simply adjust

the linear scaling according to the fraction of the z-axis distance to the focal plane at which the defocused plane lies (plus an inversion of the vector on the other side of focus). The deflection due to the phase perturbation is similarly rescaled. For typical configurations, however, the rescaling of the deflection term is very nearly one (for Rubin, the maximum deviation is only one part in ten thousand) and can therefore be safely neglected. Thus we have

$$\mathbf{x}_i^{\pm} = \mp \frac{\ell}{\sqrt{f^2 - \mathbf{x}_p^2}} \mathbf{x}_p - f \nabla_{\mathbf{x}} W(\mathbf{x}_p). \tag{B.8}$$

In the absence of wavefront perturbations, the radius of the pupil projected onto the defocused image plane is

$$R_i = \frac{\ell \, R_p}{\sqrt{f^2 - R_p^2}} = \frac{\ell}{\sqrt{4N^2 - 1}}, \tag{B.9}$$

where N is the focal ratio (or f-number) of the telescope. It is convenient to rewrite Eq. B.8 in terms of the normalized, centered coordinates

$$\mathbf{u} = (u, v) \equiv \frac{1}{R_p} \mathbf{x} \tag{B.10}$$

$$\mathbf{u}' = (u', v') \equiv \frac{1}{R_i} \Big[\mathbf{x} + f \nabla_{\mathbf{x}} W(\mathbf{0}) \Big]. \tag{B.11}$$

In these coordinates, the unaberrated pupil has a radius of 1 on both the pupil and image planes, and the aberrated pupil is centered on both the pupil and image planes. With these coordinate transformations, the mapping from pupil to image plane is

$$\mathbf{u}_i'^{\pm} = \mp \sqrt{\frac{4N^2 - 1}{4N^2 - \mathbf{u}_p^2}} \, \mathbf{u}_p - \frac{2N\sqrt{4N^2 - 1}}{\ell} \Big[\nabla_{\mathbf{u}} W(\mathbf{u}_p) - \nabla_{\mathbf{u}} W(\mathbf{0}) \Big]. \tag{B.12}$$

This formula is implemented in WEP as the "on axis" model. In the limit of $N \gg 1$, we have

$$\mathbf{u}_i'^{\pm} = \mp \mathbf{u}_p - \frac{4N^2}{\ell} \Big[\nabla_{\mathbf{u}} W(\mathbf{u}_p) - \nabla_{\mathbf{u}} W(\mathbf{0}) \Big], \tag{B.13}$$

which matches Eq.'s 11 and 12 of C. Roddier & F. Roddier [242] (note this paper only explicitly derives formulae for the intrafocal image, they do not discard the negligible adjustment

to the deflection term in Eq. B.7, and they do not center their rescaled coordinates). For Rubin $N = 1.234$, so the $N \gg 1$ assumption cannot be made but Eq. B.13 is included in WEP as the "paraxial" model for testing purposes.

Using the mapping from pupil to images planes we can predict the image intensity on the defocused planes. Let $I(\mathbf{u}_p)$ be the intensity on the pupil (which for distant point sources is typically uniform) and $I'(\mathbf{u}'_i)$ be the intensity on the image plane. From flux conservation,

$$I(\mathbf{u}_p)\, d\mathbf{u}_p = I'(\mathbf{u}'_i)\, d\mathbf{u}'_i = I'(\mathbf{u}'_i) \left| \frac{d\mathbf{u}'_i}{d\mathbf{u}_p} \right| d\mathbf{u}_p, \tag{B.14}$$

and therefore

$$I(\mathbf{u}_p) = I'(\mathbf{u}'_i) \left| \frac{d\mathbf{u}'_i}{d\mathbf{u}_p} \right|, \tag{B.15}$$

where $|d\mathbf{u}'_i/d\mathbf{u}_p|$ is the determinant of the Jacobian of the $\mathbf{u}_p \to \mathbf{u}'_i$ transformation.

The Jacobian is

$$\frac{d\mathbf{u}'^{\pm}_i}{d\mathbf{u}_p} = \sqrt{\frac{4N^2 - 1}{4N^2 - \mathbf{u}_p^2}} \left[\mp \mathbb{1}_2 + \frac{\mathbf{u}_p \otimes \mathbf{u}_p}{4N^2 - \mathbf{u}_p^2} \right] - \frac{2N\sqrt{4N^2 - 1}}{\ell} \frac{d^2 W(\mathbf{u}_p)}{d\mathbf{u}^2}. \tag{B.16}$$

In particular, $\mathbb{1}_2$ is the 2×2 identity matrix, the outer product is

$$\mathbf{u}_p \otimes \mathbf{u}_p = \begin{pmatrix} u_p^2 & u_p v_p \\ v_p u_p & v_p^2 \end{pmatrix}, \tag{B.17}$$

and $d^2 W(\mathbf{u}_p)/d\mathbf{u}^2$ is the Hessian matrix. This is the Jacobian of the "on axis" model in WEP. In the limit $N \gg 1$, this becomes

$$\frac{d\mathbf{u}'^{\pm}_i}{d\mathbf{u}_p} = \mp \mathbb{1}_2 - \frac{4N^2}{\ell} \frac{d^2 W(\mathbf{u}_p)}{d\mathbf{u}^2}, \tag{B.18}$$

the determinant of which matches Eq. 13 of C. Roddier & F. Roddier [242]. This is the Jacobian of the "paraxial" model in WEP.

Note that the on axis Eq.'s B.12 and B.16 reproduce Eq.'s 22-28 from B. Xin et al. [317] if you take the definition of m from the `CompensableImage` class in WEP version 8.

B.2.2 The Off-Axis Model

The equations above apply to sources near the optical axis, but at large field angles there are additional distortion effects these equations do not account for. For large field angles, therefore, we use a numerical model fit using `Batoid`. In this model, we represent the OPD in a Zernike series:

$$W(\mathbf{u}_p) = \sum_i \alpha_i Z_i(\mathbf{u}_p), \quad \big[\text{OPD Only}\big] \tag{B.19}$$

where the α_i are coefficients in meters (see Section B.3 for more details on Zernike polynomials). A second series of coefficients fit by `Batoid`, β_i, are added to these coefficients such that

$$W(\mathbf{u}_p) = \sum_i (\alpha_i + \beta_i) Z_i(\mathbf{u}_p). \quad \big[\text{OPD} + \texttt{Batoid}\ \text{model}\big] \tag{B.20}$$

We then use the equations above, dispensing with any terms that do not directly reference $W(\mathbf{u}_p)$. That is,

$$\mathbf{u}_i'^{\pm} = -\frac{2N\sqrt{4N^2-1}}{\ell}\Big[\nabla_{\mathbf{u}} W(\mathbf{u}_p) - \nabla_{\mathbf{u}} W(\mathbf{0})\Big] \quad \text{and} \quad \frac{d\mathbf{u}_i'^{\pm}}{d\mathbf{u}_p} = -\frac{2N\sqrt{4N^2-1}}{\ell}\frac{d^2 W(\mathbf{u}_p)}{d\mathbf{u}^2}. \tag{B.21}$$

Note the Batoid coefficients only account for off-axis projection effects and do not account for the intrinsic aberrations associated with the telescope. These formulae are implemented in WEP as the "off axis" model.

B.3 Inferring the Wavefront from Defocused Images

The goal of the active optics system is to infer $W(\mathbf{u}_p)$, the OPD of the optical system, from the defocused image intensity $I_i(\mathbf{u}_i, z_i = f \pm \ell)$ using the relations derived in the previous section. This is an inversion problem. In the following subsections, I detail three different methods for performing this inversion.

In all cases, we seek to represent the OPD as a linear combination of annular Zernike polynomials defined on the pupil:

$$W(\mathbf{u}_p) = \sum_i \alpha_i Z_i(\mathbf{u}_p), \tag{B.22}$$

where the coefficients α_i carry the units. The annular Zernike polynomials are a complete orthogonal set of basis functions that roughly correspond to traditional optical aberrations. We use the Noll index scheme – i.e. Z_4 corresponds to defocus, Z_5 and Z_6 correspond to oblique and vertical astigmatism, etc. We use the normalization

$$\iint Z_i(\mathbf{u}_p) Z_j(\mathbf{u}_p)\, d\mathbf{u}_p = A\delta_{ij}, \tag{B.23}$$

where A is the pupil area. Using the normalized coordinates defined in the previous section, $A = \pi(1-\varepsilon^2)$ where ε is the fractional obscuration of the pupil. For Rubin this is $\varepsilon = 0.612$.

B.3.1 Inversion via forward modeling

One method for estimating the Zernike coefficients is to guess an initial set of coefficients and use the formulae from the previous section to forward model the image on one (or both) of the defocused plane(s). You can then use an optimization routine to determine the set of Zernike coefficients that best matched the observed image. This is the method proposed by R. J. Janish [152] and implemented in the `Danish` algorithm.

In addition to optimizing the Zernike coefficients, `danish` also jointly optimizes a source pixel offset and the astronomical seeing. The former allows `danish` to account for miscentering, while the latter accounts for the atmospheric turbulence that blurs any image taken with a ground based telescope.

`danish` also implements a mask model to account for vignetting which it also projects onto the image plane. This is similar to the mask model from R. J. Janish [152] which estimates the fractional illumination of pixels on the pupil edge by making a locally linear approximation of the circular pupil across each of the edge pixels. However the projection of the pupil onto the image plane is only circular in the absence of wavefront aberrations.

`danish` improves on the model of R. J. Janish [152], therefore, by accounting for wavefront aberrations in the mask model.

Early testing suggests this method is prone to outliers, but is otherwise more accurate than the TIE (described below). `danish` is also slower than the TIE.

B.3.2 Inversion via Solving the Transport of Intensity Equation

If you have a pair of images from both sides of focus, you can estimate the wavefront by solving the transport of intensity equation (TIE). The advantage of the TIE over the forward modeling method is that it admits an algebraic solution which is faster than general optimization. In reality the TIE must be iteratively solved to converge on the true wavefront, however even with these iterations the TIE method is typically faster than the forward modeling method. The disadvantage is that the TIE method is complicated and more sensitive to issues such as miscentering, blending, and vignetting.

We start by deriving the TIE in Section B.3.2. In Section B.3.2 we discuss approximations of the beam intensity and longitudinal derivative, which are key inputs to the TIE. Finally, in Section B.3.2 we discuss solving the TIE via Zernike expansion.

Deriving the TIE

We will derive the TIE by applying conservation of energy in the paraxial limit. You can also derive the TIE as the imaginary part of the paraxial Helmholtz equation (i.e. the time-independent wave equation). Note these paraxial assumptions concern the incoming photon beam, which for distant point sources is very nearly a plane wave. This differs from the previous section in which we considered a photon beam after it had passed through the optical system. In that case, the paraxial assumption is invalid for systems with low focal ratio, like the Rubin Observatory's Simonyi Survey Telescope.

The complex amplitude of a monochromatic wave with intensity I and phase φ can be

written

$$\Psi(z,\mathbf{x}) = \sqrt{I(z,\mathbf{x})}e^{i\varphi(z,\mathbf{x})}. \tag{B.24}$$

The phase of this wave can be expanded

$$\varphi(z,\mathbf{x}) = k_z z + \mathbf{k_x}\cdot\mathbf{x}, \tag{B.25}$$

where

$$k^2 = k_z^2 + |\mathbf{k_x}|^2 = \left(\frac{2\pi}{\lambda}\right)^2, \tag{B.26}$$

and λ is the wavelength of the light.

The Poynting vector is the energy flux of the wave. By definition, the intensity is the magnitude of the time average of the Poynting vector: $|\langle\mathbf{S}(z,\mathbf{x})\rangle| = I(z,\mathbf{x})$. Since photons propagate perpendicular to the wavefront, the time average Poynting vector must also be proportional to the gradient of the phase: $\langle\mathbf{S}(z,\mathbf{x})\rangle \propto \nabla\varphi(z,\mathbf{x})$. Combining these two requirements, we have

$$\langle\mathbf{S}(z,\mathbf{x})\rangle = \frac{1}{k}I(z,\mathbf{x})\nabla\varphi(z,\mathbf{x}). \tag{B.27}$$

Now let's use the paraxial approximation. That is, we consider a wave whose direction is very nearly aligned with the z axis. In this limit, $|\mathbf{k}_{x,y}|/k_z \ll 1$, and we can approximate $k_z \approx k$. Then,

$$\nabla\varphi(z,\mathbf{x}) = k\hat{\mathbf{z}} + \nabla_{\mathbf{x}}\varphi(\mathbf{x}), \tag{B.28}$$

where $\nabla_{\mathbf{x}}$ is the transverse gradient and $\varphi(\mathbf{x})$ is the transverse component of the phase. We can now write

$$\langle\mathbf{S}(\mathbf{x},z)\rangle = I(\mathbf{x},z)\hat{\mathbf{z}} + \frac{1}{k}I(\mathbf{x},z)\nabla_{\mathbf{x}}\varphi(\mathbf{x}). \tag{B.29}$$

Finally, we can apply conservation of energy. In the vacuum, energy conservation demands that

$$\nabla\cdot\langle\mathbf{S}(z,\mathbf{x})\rangle = 0. \tag{B.30}$$

Plugging in the time-averaged Poynting vector in the paraxial limit, we get the TIE:

$$\frac{\partial I(z,\mathbf{x})}{\partial z} + \nabla_{\mathbf{x}} \cdot \left[\frac{1}{k} I(z,\mathbf{x}) \nabla_{\mathbf{x}} \varphi(\mathbf{x}) \right] = 0. \tag{B.31}$$

The TIE has a simple interpretation in terms of conservation of energy. In the paraxial limit, $I(z,\mathbf{x})$ is the energy flux in the longitudinal direction, and $k^{-1} I(z,\mathbf{x}) \nabla_{\mathbf{x}} \varphi(\mathbf{x})$ is the energy flux in the transverse direction. The TIE says the decrease (increase) in longitudinal energy flux must be equal to the energy flux that is leaking out (coming in) the transverse direction.

We can see another obvious interpretation of the TIE if we use the fact that the pupil is uniformly illuminated by distant point sources — i.e., $\nabla_{\mathbf{x}} I = 0$ inside the pupil. Using this fact and applying our definition of the OPD yields

$$\nabla^2 W = \frac{1}{I} \frac{\partial I}{\partial z}. \tag{B.32}$$

In words, fractional changes in beam intensity are sourced by wavefront curvature.

Finally, we can rewrite Eq. B.31 in the normalized pupil coordinates of Section B.2:

$$\frac{\partial I(0,\mathbf{u}_p)}{\partial z} = \frac{1}{R^2} \nabla_{\mathbf{u}} \cdot [I(0,\mathbf{u}_p) \nabla_{\mathbf{u}} W(\mathbf{u}_p)]. \tag{B.33}$$

Approximating Beam Intensity and Longitudinal Derivative

To solve the TIE, we imagine that our defocused images are in-focus images of different cross sections of the photon beam. We map the defocused images back to the pupil plane, but rather than thinking of the result as pupil illuminations, we imagine these are physical cross sections of the photon beam. We can use the lens equation to determine the z positions of these cross sections:

$$\frac{1}{o} + \frac{1}{f} = \frac{1}{i}, \tag{B.34}$$

where o and i are the object and image positions, respectively. We know our images are at locations $i = f \pm \ell$, so the objects are at locations

$$o = \mp \frac{f}{\ell} (f \pm \ell). \tag{B.35}$$

Thus, we see that the extrafocal image ($i = f + \ell$) correspond to a cross section of the beam before the exit pupil ($o < 0$), while the intrafocal image ($i = f - \ell$) correspond to a cross section after the exit pupil ($o > 0$). For Rubin, $f \gg \ell$, and so we can state that $o \approx \mp \Delta z$, where

$$\Delta z = \frac{f^2}{\ell}. \tag{B.36}$$

With this knowledge in hand, we can make a linear approximation of I and $\partial I/\partial z$:

$$I(0, \mathbf{u}_p) \approx \frac{I(+\Delta z, \mathbf{u}_p) + I(-\Delta z, \mathbf{u}_p)}{2} \tag{B.37}$$

$$\frac{\partial I(0, \mathbf{u}_p)}{\partial z} \approx \frac{I(+\Delta z, \mathbf{u}_p) - I(-\Delta z, \mathbf{u}_p)}{2\Delta z}. \tag{B.38}$$

Note this approximation is valid for small Δz. Counter-intuitively, small values of Δz correspond to large values of ℓ. For Rubin ℓ is quite small (meaning $\ell \ll f$) which violates this assumption. Regardless, C. Roddier & F. Roddier [242] found that solving the TIE in the iterative manner described below converges to an accurate estimate of the wavefront.

Solving via direct Zernike expansion

With approximations of I and $\partial I/\partial z$ in hand, we need to invert the TIE and solve for the OPD. This can be done by expanding the OPD in a Zernike series:

$$\frac{\partial I(0, \mathbf{u}_p)}{\partial z} = \frac{1}{R^2} \sum_j \alpha_j \nabla_{\mathbf{u}} \cdot [I(0, \mathbf{u}_p) \nabla_{\mathbf{u}} Z_j(\mathbf{u}_p)]. \tag{B.39}$$

We can then multiply each side by $Z_i(\mathbf{u}_p)$ and integrate over the pupil:

$$\iint Z_i(\mathbf{u}_p) \frac{\partial I(0, \mathbf{u}_p)}{\partial z} \, d\mathbf{u}_p = \frac{1}{R^2} \sum_j \alpha_j \iint Z_i(\mathbf{u}_p) \nabla_{\mathbf{u}} \cdot [I(0, \mathbf{u}_p) \nabla_{\mathbf{u}} Z_j(\mathbf{u}_p)] \, d\mathbf{u}_p. \tag{B.40}$$

Integrating by parts and using the fact that the beam intensity vanishes at the boundary yields

$$\iint Z_i(\mathbf{u}_p) \frac{\partial I(0, \mathbf{u}_p)}{\partial z} \, d\mathbf{u}_p = -\frac{1}{R^2} \sum_j \alpha_j \iint I(0, \mathbf{u}_p) \, \nabla_{\mathbf{u}} Z_i(\mathbf{u}_p) \cdot \nabla_{\mathbf{u}} Z_j(\mathbf{u}_p) \, d\mathbf{u}_p. \tag{B.41}$$

If we discretize the pupil into pixels indexed by a, b, the integrals become summations and the pixel size cancels, yielding

$$\sum_{a,b} (Z_i)_{ab} \left(\frac{\partial I}{\partial z}\right)_{ab} = \sum_j \alpha_j \left[-\frac{1}{R^2} \sum_{ab} I_{ab} \left(\nabla_{\mathbf{u}} Z_i \cdot \nabla_{\mathbf{u}} Z_j\right)_{ab}\right]. \tag{B.42}$$

This is just a linear system. If we define the vector

$$b_i = \sum_{a,b} (Z_i)_{ab} \left(\frac{\partial I}{\partial z}\right)_{ab} \tag{B.43}$$

and the matrix

$$M_{ij} = -\frac{1}{R^2} \sum_{a,b} I_{ab} \left(\nabla_{\mathbf{u}} Z_i \cdot \nabla_{\mathbf{u}} Z_j\right)_{ab}, \tag{B.44}$$

then we seek to solve the equation

$$b_i = \sum_j M_{ij} \alpha_j. \tag{B.45}$$

Of course with real data there is no guarantee a solution exists. Instead, we treat this as a regression problem and minimize the squared residuals of b_i. In WEP this is done using `np.linalg.lstsq`.

Since both b_i and M_{ij} are constructed from the same pair of images, both are noisy and have correlated errors. Therefore, you might expect an errors-in-variables model is needed for unbiased estimation of the α_i. However, a simple argument shows that the noise in M_{ij} can be neglected and ordinary least squares is appropriate:

1. The errors in I and $\partial I/\partial z$ are of the same order of magnitude, and R^2 is $\mathcal{O}(10^2\,\mathrm{m})$.

2. Due to the normalization condition (Eq. B.23), the values of Z_i and $\nabla_{\mathbf{u}} Z_i$ are $\mathcal{O}(1)$.

3. The errors in M_{ij} are connected to the variables b_i by the Zernike coefficients α_j, which are $\alpha_j \lesssim \mathcal{O}(10^{-6}\,\mathrm{m})$ by assumption.

4. Therefore the errors in M_{ij} propagated to b_i are roughly 8 orders of magnitude smaller than the errors native to b_i.

Furthermore, the wavefront error due to pixel noise is subdominant to the phase variance due to atmospheric turbulence. For the impact of atmospheric turbulence on Zernike coefficient estimation, see Appendix D of J. F. Crenshaw et al. [71].

B.3.3 Inversion via Artificial Intelligence

Rather than implementing the mapping equations of Section B.2 to solve for the wavefront, one can build a deep learning (or artificial intelligence; AI) network to implicitly learn the mapping by training the network to estimate Zernike coefficients directly from out-of-focus images. For more details on this approach and comparisons to the TIE see J. F. Crenshaw et al. [71].

Appendix B.a: Pupil to Image Mapping for Off-axis Points

In this appendix we will generalize the derivations of Section B.2 for off-axis image points.

The wavefront of a spherical wave emerging from the pupil and converging on a point $(z, \mathbf{x}) = (f, \mathbf{x}_i^0)$ is

$$\varphi(z, \mathbf{x}) = -k\sqrt{(z-f)^2 + (\mathbf{x} - \mathbf{x}_i^0)^2}. \tag{B.46}$$

The phase gradient on the reference sphere is

$$\nabla\varphi(\mathbf{x}_p) = \frac{k}{\tilde{f}}\left(\sqrt{\tilde{f} - (\mathbf{x}_p - \mathbf{x}_i^0)^2}\;\hat{\mathbf{z}} - (\mathbf{x}_p - \mathbf{x}_i^0)\right), \tag{B.47}$$

where we defined $\tilde{f} = \sqrt{f^2 + (\mathbf{x}_i^0)^2}$ and used $z_p = f - \sqrt{\tilde{f}^2 - (\mathbf{x}_p - \mathbf{x}_i^0)^2}$. You can now see the unperturbed wave emerges from $(z_p, \mathbf{x}_p)$ and converges on $(f, \mathbf{x}_i^0)$. Again adding the phase perturbation $\delta\varphi = -k\,W(\mathbf{x}_p)$ yields the gradient

$$\nabla\varphi(\mathbf{x}_p) = \frac{k}{\tilde{f}}\left(\sqrt{\tilde{f} - (\mathbf{x}_p - \mathbf{x}_i^0)^2}\;\hat{\mathbf{z}} - (\mathbf{x}_p - \mathbf{x}_i^0) - \tilde{f}\nabla_{\mathbf{x}} W(\mathbf{x}_p)\right). \tag{B.48}$$

Using the same similar-triangles argument, we have

$$\frac{\mathbf{x}_i^{\pm} - \mathbf{x}_p}{z_i^{\pm} - z_p} = \frac{\nabla_{\mathbf{x}}\varphi(\mathbf{x}_p)}{\partial\varphi(\mathbf{x}_p)/\partial z} \implies \frac{\mathbf{x}_i^{\pm} - \mathbf{x}_p}{\sqrt{f^2 - \mathbf{x}_p} \pm \ell} = \frac{-(\mathbf{x}_p - \mathbf{x}_i^0) - \tilde{f}\nabla_{\mathbf{x}}W(\mathbf{x}_p)}{\sqrt{\tilde{f} - (\mathbf{x}_p - \mathbf{x}_i^0)^2}}. \tag{B.49}$$

Isolating the image position:

$$\mathbf{x}_i^{\pm} - \mathbf{x}_i^0 = \mp\frac{\ell}{\sqrt{\tilde{f}^2 - (\mathbf{x}_p - \mathbf{x}_i^0)^2}}\,\mathbf{x}_p - \tilde{f}\left(1 \pm \frac{\ell}{\sqrt{\tilde{f}^2 - (\mathbf{x}_p - \mathbf{x}_i^0)^2}}\right)\nabla_{\mathbf{x}}W(\mathbf{x}_p). \tag{B.50}$$

Again, for the aberration term we can safely discard the term of order $\mathcal{O}(\ell/\tilde{f})$, yielding

$$\mathbf{x}_i^{\pm} - \mathbf{x}_i^0 = \mp\frac{\ell}{\sqrt{\tilde{f}^2 - (\mathbf{x}_p - \mathbf{x}_i^0)^2}}\,\mathbf{x}_p - \tilde{f}\nabla_{\mathbf{x}}W(\mathbf{x}_p). \tag{B.51}$$

We can use the same normalized coordinates of Eq. B.11, except we now absorb $\mathbf{x}_i^0$ into the centering of the image coordinates. We can also use $\mathbf{x}_i^0 = -f\theta = -2NR_p\,\theta$ where the θ is the field angle of the source in radians, and the negative sign is due to the fact that Rubin has an odd number of mirrors (for Rubin this formula is accurate to one part in ten thousand even at wide field angles). With these definitions, and plugging in the definition of $\tilde{f}$, Eq. B.51 becomes

$$\mathbf{u}_i'^{\pm} = \mp\sqrt{\frac{4N^2 - 1}{4N^2 - \mathbf{u}_p^2 + 4N(\mathbf{u}_p \cdot \theta\,)}}\;\mathbf{u}_p - \frac{2N\sqrt{(4N^2-1)(1-\theta^2)}}{\ell}\Big[\nabla_{\mathbf{u}}W(\mathbf{u}_p) - \nabla_{\mathbf{u}}W(\mathbf{0})\Big]. \tag{B.52}$$

Taking the derivative of this mapping yields the Jacobian

$$\frac{d\mathbf{u}_i'^{\pm}}{d\mathbf{u}_p} = \sqrt{\frac{4N^2 - 1}{4N^2 - \mathbf{u}_p^2 + 4N(\mathbf{u}_p \cdot \theta\,)}}\Big[\mp\mathbb{1}_2 + \frac{(\mathbf{u}_p - 2N\theta) \otimes \mathbf{u}_p}{4N^2 - \mathbf{u}_p^2 + 4N(\mathbf{u}_p \cdot \theta\,)}\Big] - \frac{2N\sqrt{(4N^2-1)(1-\theta^2)}}{\ell}\frac{d^2W(\mathbf{u}_p)}{d\mathbf{u}^2}. \tag{B.53}$$

These formulae mildly improve the accuracy of the on axis model when applied to off-axis points near the center of the focal plane. However, these formulae plus the inclusion of the intrinsic telescope aberrations in $W(\mathbf{u}_p)$ are not sufficient for accurately modeling wide field angles for Rubin. As such, these formulae are not currently implemented in WEP and the numerical off-axis model described in Section B.2.2 is used instead.

Appendix C

USING AI FOR WAVEFRONT ESTIMATION WITH THE RUBIN OBSERVATORY ACTIVE OPTICS SYSTEM

This chapter was originally published in *The Astronomical Journal 167, 86 (2024).*

C.1 Abstract

The Vera C. Rubin Observatory will, over a period of 10 years, repeatedly survey the southern sky. To ensure that images generated by Rubin meet the quality requirements for precision science, the observatory will use an Active Optics System (AOS) to correct for alignment and mirror surface perturbations introduced by gravity and temperature gradients in the optical system. To accomplish this Rubin will use out-of-focus images from sensors located at the edge of the focal plane to learn and correct for perturbations to the wavefront. We have designed and integrated a deep learning model for wavefront estimation into the AOS pipeline. In this chapter, we compare the performance of this deep learning approach to Rubin's baseline algorithm when applied to images from two different simulations of the Rubin optical system. We show the deep learning approach is faster and more accurate, achieving the atmospheric error floor both for high-quality images, and low-quality images with heavy blending and vignetting. Compared to the baseline algorithm, the deep learning model is 40x faster, the median error 2x better under ideal conditions, 5x better in the presence of vignetting by the Rubin camera, and 14x better in the presence of blending in crowded fields. In addition, the deep learning model surpasses the required optical quality in simulations of the AOS closed loop. This system promises to increase the survey area useful for precision science by up to 8%. We discuss how this system might be deployed when commissioning and operating Rubin.

C.2 Introduction

The Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) will spend ten years imaging the entire southern sky to unprecedented depth [146]. LSST promises to improve cosmological constraints by an order of magnitude [291], dramatically expand our understanding of galaxy evolution [241], unveil the transient and variable sky [126], and provide an inventory of the Solar System of unprecedented completeness [271].

To enable the wide variety of science undertaken by LSST, the Rubin Observatory must deliver high and consistent optical quality across the entire 3.5 degree field of view (FoV). The median seeing at Rubin's site on Cerro Panchón is 0.65" [146], and Rubin's optical system is required to degrade this value by no more than 0.4" (see Appendix C.6). To meet this requirement, the Rubin Observatory will use an active optics system (AOS) to provide real-time corrections to the optical alignment and mirror figure, which are perturbed by gravity and temperature gradients [217, 295]. The AOS uses 228 actuators to control the surface figure of the mirrors (156 actuators on the joint primary/tertiary mirror, called M1M3, and 72 actuators on the secondary mirror, called M2), and 2 hexapods to adjust the positions and rotations of M2 and the camera. A position- and temperature-dependent lookup table (LUT) is used in an open-loop to apply optimal mean corrections before each exposure. These mean corrections are, however, unable to account for hysteresis, dome seeing, small-scale temperature gradients, and temperature uncertainties.

The Rubin AOS will use a closed-loop to detect and correct these residual errors. At each corner of the focal plane, there are curvature wavefront sensors (CWFSs) consisting of an offset pair of CCDs: one 1.5 mm inside of focus (intra-focal) and one 1.5 mm outside of focus (extra-focal), corresponding to 24 μm RMS of wavefront defocus. These out-of-focus images are used to infer perturbations to the optical wavefront, from which actuator and hexapod corrections are determined. Due to the fast cadence of LSST, wavefront inference and correction of the optical system must be completed in less than 12 seconds, and this whole process must be repeated every 36 seconds for the duration of an observing run. To

deliver the optical quality required for LSST, the AOS must limit the PSF (Point Spread Function) FWHM (Full Width Half Maximum) contribution of optical aberrations to no more than 0.09", including only 0.079" due to misestimation of the wavefront – less than 10% of the total error budget (see Appendix C.6).

There are several features of Rubin's design that make wavefront estimation difficult: (i) the primary mirror has a large central obscuration ($R_{\text{inner}}/R_{\text{outer}} = 0.61$), which results in a large loss of information about the wavefront; (ii) the optical beam is very fast ($f/1.23$), which results in a non-linear projection of the reference sphere onto the pupil; (iii) the CWFSs are at the edge of Rubin's wide field of view, generating pupil distortion and vignetting of the out-of-focus images; (iv) the images from the intra- and extra-focal sensors are of different sources, which have different intrinsic properties and field-dependent wavefront variations that must be accounted for.

The baseline algorithm for wavefront inference is an iterative solver of the transport of intensity equation (TIE), described in B. Xin et al. [317]. This algorithm has been implemented and tested in the Rubin AOS pipeline [296], but there are several limitations that could be improved upon, including degradation of wavefront estimation in the presence of vignetting from the camera body and blending in crowded fields, and slow evaluation times. The vignetting and blending limitations constrain source selection for wavefront estimation, and in the densest 8% of fields it is expected that there will be no sources available for reliable wavefront estimation with the TIE. The limitation on time reduces the number of sources that can be used to constrain the wavefront.

In this chapter, we build, integrate, and test a deep learning (DL) wavefront estimator that uses convolutional neural nets (CNNs) to learn a mapping from individual out-of-focus images and metadata to the optical wavefront. Previous papers have built prototype CNNs for wavefront estimation with Rubin [292, 293], but this is the first time such a system has been integrated into the Rubin AOS pipeline, and validated against the baseline algorithm. In addition, J. E. Yin et al. [319] built a prototype network that directly estimates control parameters, without the need for first estimating the wavefront. In this chapter, we focus on

estimating the wavefront, as this allows direct comparison to the baseline Rubin algorithm. We show that our system outperforms the baseline algorithm on simulations of Rubin images, including dramatic improvements in the case of vignetting and blending, and faster evaluation times.

In Section C.3 we describe the simulations we use to train and test our system. In Section C.4 we describe the baseline wavefront estimation algorithm, and the details of our deep learning approach. In Section C.5 we compare the performance of our deep learning system to the baseline algorithm. We discuss these results and conclude in Section C.6.

C.3 Simulated data

In this section, we describe the simulations used to train and test the deep learning model. In particular, we train the network on 256 000 donuts simulated with the Batoid [204] package, and simulate validation and test sets each containing 32 000 donuts. The validation set is used to avoid overfitting during training, and the test set is used for computing metrics for comparison to the baseline wavefront estimation algorithm. We also simulate the AOS closed loop using an independent raytracing package PhoSim [230], which allows us to test whether the network has overfit on details specific to the Batoid simulation.

C.3.1 Batoid simulations

The network training data and the test data for the metrics in Section C.5.1 were generated using the Batoid [204] raytracer to simulate the Rubin optical system. The atmosphere, photometric bandpasses, and image sensors were simulated using GalSim [245]. The components of these simulations are detailed below:

1. We select a set of observing conditions from the baseline LSST OpSim simulation [191]. This includes the photometric band, the airmass, the 500 nm seeing at zenith, and the sky brightness. For details on the simulated atmosphere, see Appendix C.6.

2. From the baseline OpSim simulation, we select a random astronomical pointing, and

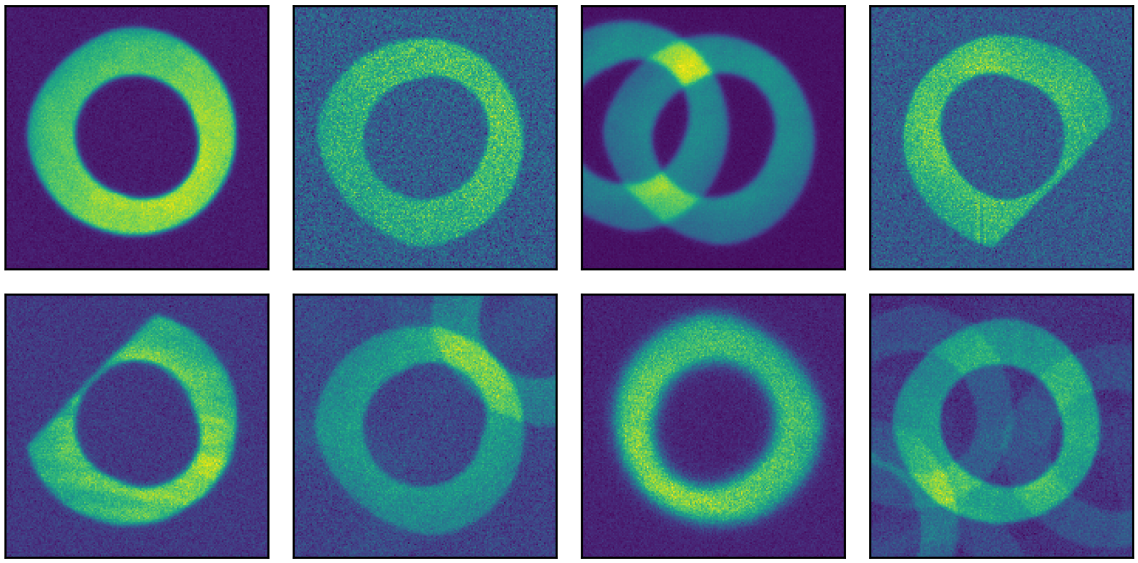

Figure C.1: Example of donuts simulated with Batoid and Galsim. This set represents the diversity of the simulation: the simulated donuts have varying SNR, some are sharply vignetted by the Rubin camera (top right and bottom left), some are blended with a single other donut, and a small fraction are blended with many donuts.

query the Gaia [111] DR2 [112] reference catalog for stars at this pointing that fall on the CWFSs. For each star, we retrieve the position, and the apparent Gaia magnitudes (G, G_{BP}, G_{RP}). We use the following equation from Chapter 5 of the Gaia DR2 documentation [41] to estimate the equivalent magnitude in the SDSS r band:

$$\begin{aligned} r = G + 0.12879 - 0.24662 \cdot x \\ + 0.027464 \cdot x^2 + 0.049465 \cdot x^3, \end{aligned} \tag{C.1}$$

where $x = G_{BP} - G_{RP}$. We then assume the LSST r magnitude is the same as SDSS, sample a random temperature in the range $4 \cdot 10^3 K < T < 10^4 K$, and use a blackbody spectrum to calculate the observed magnitude in the given LSST band. We clip bright magnitudes at 14 to save computational time during photon raytracing.

3. We generate a random set of perturbations to the Rubin optical state, and extract the Zernike coefficients of the corresponding wavefront perturbation. This is detailed in Appendix C.6.

4. Given the set of stars and the estimated sky background we simulate these sources using Batoid and Galsim, with an exposure time of 15 seconds.

5. We select the 8 brightest sources per CWFS and cut out stamps of 160x160 pixels, yielding 32 donuts per simulation.

The slowest step is the atmosphere generation, so for each atmosphere generated in step 1, we repeat steps two through five 100 times. We generate 100 different atmospheres, resulting in a total of 320 000 donut stamps, which we split 80%/10%/10% into training/validation/test sets. Each of these sets have distinct sets of atmospheres and optical perturbations.

Fig. C.1 shows eight example donut stamps. The shape of the donuts is due to their position at the far edge of the field of view, resulting in pupil distortion and vignetting of rays by the telescope and camera bodies. For example, the left two donuts in the top row have soft vignetting from the edges of the telescope mirrors, while the top right and bottom

left donuts have sharp vignetting from the camera body. Some of these donuts have bright blends, which we define as an overlapping donut within 2 magnitudes of the central donut. Others have faint blends. The impact of these donut features on wavefront estimation will be explored in Section C.5.1.

C.3.2 PhoSim

To provide a test of the robustness of the neural network to the properties of the training sample (prior to real data being obtained by Rubin), we evaluate our Batoid-trained network on simulations from PhoSim [230], an independent raytracing simulation. PhoSim simulates the atmosphere, telescope, camera, and detector, and was developed for the Rubin Observatory to model the expected images from the telescope. PhoSim uses different simulation strategies from Batoid and GalSim, and produces donut images that are not identical to the Batoid donuts, both in their shapes and intensity patterns. In addition, the PhoSim simulations we use have some features, such as mirror print through, that are not present in our Batoid simulations. These features make the PhoSim simulations a good candidate for testing how well our model adapts to other data sets. In Section C.5.2, we use our deep learning model to estimate wavefronts as a part of the AOS closed loop simulated with PhoSim.

C.4 Wavefront Estimation

The Rubin AOS uses out-of-focus images of stars, which appear as donuts, to estimate the wavefront of the telescope. The wavefront, φ, is expressed as a linear combination of annular Zernike polynomials [193], Z_i, with expansion coefficients

$$\alpha_i = \frac{1}{A} \int d^2\rho \; Z_i(\rho)\varphi(\rho). \tag{C.2}$$

A is the area of the pupil, which implicitly defines the normalization of Z_i, and ρ is a 2D position vector on the pupil. The goal of wavefront estimation is to estimate the values of

the coefficients α_i. Only coefficients 4-22 (numbered according to Noll's convention [222]) are currently used to define perturbations in Rubin's optical system [317].

Note that even with perfect alignment and mirror figure, every telescope has non-zero off-axis optical aberrations, i.e. $\alpha_i^0 \neq 0$. Since we do not aim to "correct" these intrinsic aberrations from the design of the telescope, we seek to instead estimate $\alpha_i - \alpha_i^0$. We will continue using the symbol α_i and the term "wavefront", leaving the subtraction of the intrinsic aberrations implicit. Furthermore, as the wavefront varies across the focal plane, we aim to estimate the wavefront at the center of the boundary between the intra- and extra-focal chip of each CWFS pair. After estimating the wavefront, we scale each coefficient α_i by a factor s_i, such that $|s_i \alpha_i|$ is approximately the degradation of the PSF FWHM due to the excitation of Zernike mode Z_i (see Appendix C.6 for details). For the rest of this chapter, we use this scaling to express Zernike amplitudes and wavefront estimation errors in terms of the equivalent PSF FWHM degradation.

The baseline wavefront estimation algorithm of the Rubin AOS is explained in Section C.4.1. In Section C.4.2, we discuss our deep learning model, including the design and training of the neural network.

C.4.1 Baseline wavefront estimation with the transport of intensity equation

The baseline wavefront estimation algorithm of the Rubin AOS uses the transport of intensity equation (TIE),

$$\nabla_\rho \cdot [I(\rho, 0) \nabla_\rho \varphi(\rho)] + k \frac{\partial I(\rho, z)}{\partial z}|_{z=0} = 0. \tag{C.3}$$

$I(\rho, 0)$ is the intensity pattern on the pupil, which is located at $z = 0$. The TIE is essentially a conservation of energy equation in the paraxial limit, and relates the z-derivative of the intensity to the curvature of the wavefront. The AOS uses a pair of intra- and extra-focal images to approximate the intensity at the pupil and the z-derivative across the pupil. The

TIE is then solved using an expansion of the wavefront in annular Zernike polynomials[1].

The Rubin AOS does not take images of the same stars on both sides of focus, but rather relies on simultaneous images of different stars on distinct intra- and extra-focal sensors. This formally violates the assumptions of the TIE, and requires additional steps to mitigate problems including different pupil masks and intensities, and variations of the wavefront across the focal plane. The fast f-number, and the location of the CWFSs at the very edge of the field of view add additional complications. B. Xin et al. [317] explains the algorithm in detail. The existing implementation of this TIE solver in the Rubin AOS pipeline will serve as the baseline against which our deep learning model is compared.

C.4.2 Deep learning wavefront estimation

The deep learning (DL) wavefront estimator is summarized in Fig. C.2. A convolutional neural network (CNN) extracts information from a single donut image, and produces a set of image features. These features, along with some metadata, are fed into a set of fully connected layers that estimate the Zernike coefficients of the wavefront. The metadata consists of: (i) a flag indicating whether the image is intra- or extra-focal; (ii) the field angle of the star; (iii) the effective wavelength of the photometric band in which the donut was imaged.

We test two different architectures, both of which have the structure presented in Fig. C.2, but differ in the details of the network layers contained inside the CNN and fully connected stages. The first is the architecture designed for wavefront estimation by D. Thomas et al. [293]. The second uses ResNet-18 [131] (with the final prediction layer removed) as the CNN, plus three fully connected layers to predict Zernike coefficients from the image features and metadata. ResNet-18 is a network that was trained on image classification for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). We initialize ResNet-18 with its pre-trained parameters, and test versions where those parameters are frozen during training, and

[1]The solution can be found either by projecting onto the annular Zernike basis, or via Fast Fourier Transform (FFT). For this work, we use the projection method.

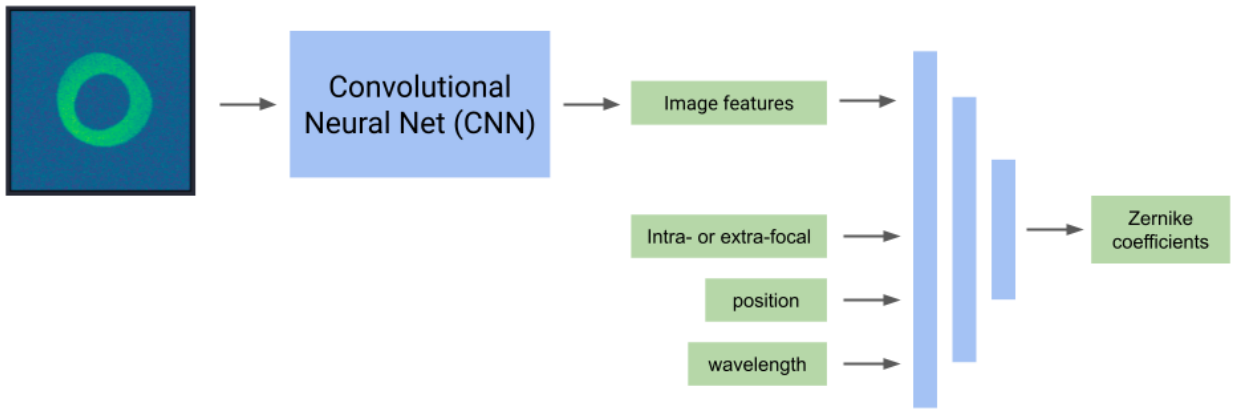

Figure C.2: Schematic of the network architecture. A single donut is passed through a convolutional neural network (CNN), generating a set of image features. These features are concatenated with a flag that indicates if the donut is intra-/extra-focal, the field angle of the source, and the effective wavelength of the observation. A dense neural network then estimates the Zernike coefficients associated with the optical wavefront.

where those parameters are free to change with the rest of the network parameters. For the fully connected layers, we use three layers of size (171, 57, 19), the first two of which are followed by batch normalization and ReLU activation. The output dimension 19 corresponds to the 19 Zernike coefficients we are estimating. The former two dimensions were chosen to be equally spaced logarithmically between 516 (512 image features output by ResNet-18 + 4 items of metadata) and the output dimension (19).

For both network architectures, we apply the same preprocessing to the data. We rescale each image so that all pixel values are between 0 and 1. We then calculate the mean and variance of all pixels in the training images, and whiten the pixels

$$\text{pixel} \to \frac{\text{pixel} - \text{mean}}{\sqrt{\text{variance}}}. \tag{C.4}$$

We also calculated the training-set mean and variance of each item of metadata, and whitened them as well. Note that these same transformations are applied to the test set before network

evaluation, using the means and variances calculated from the training set.

We train the networks to minimize the residual sum of squares (RSS) of the estimated Zernike coefficients and the true coefficients, including the weighting for the PSF FWHM degradation. In other words, the network is trained to minimize the PSF width. We begin training using Adam [160] with a learning rate of 3×10^{-4}, and decrease the learning rate by a factor of 10 whenever the validation loss does not decrease for 10 epochs. We stop training whenever the validation loss has not improved for 20 epochs.

When training the network whose CNN is the pre-trained ResNet-18, we have a choice to make about how to handle the pre-trained parameters. We can keep the pre-trained parameters fixed during training, and only allow the fully connected layers to adjust their parameters, or we can unfreeze ResNet-18, and allow its parameters to be tweaked during training as well. We found that unfreezing the ResNet-18 parameters resulted in a much better validation loss.

Both the network from D. Thomas et al. [293] and ResNet-18 with unfrozen parameters achieved good validation losses, plateauing near the atmospheric error floor (see Section C.5.1 and Appendix C.6). However, the network from D. Thomas et al. [293] did not achieve as high of an optical quality as the unfrozen ResNet-18 when tested in the simulated AOS closed loop (see Section C.5.2). This suggests it is not as adept at domain adaptation as ResNet-18. For the rest of this chapter, we consider only the model that uses ResNet-18 with unfrozen parameters. We note that J. E. Yin et al. [319] tested a wider variety of neural network architectures for a similar AOS task, and also found that ResNet-18 achieved the best results.

C.5 Results

In this section we compare the performance of the deep learning (DL) model to the baseline wavefront estimator (TIE). First, we evaluate performance on the Batoid test set. Then, we test both algorithms on the AOS closed loop simulated with PhoSim. For both of these sections, we use the DL algorithm to estimate the wavefront from pairs of intra- and extra-

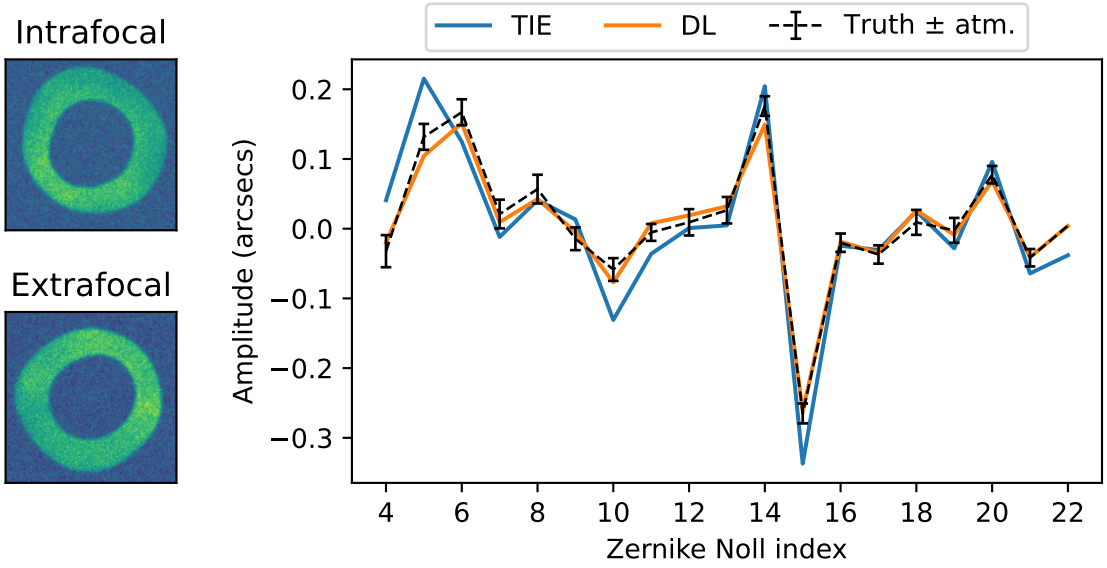

Figure C.3: Example of an intra- and extra-focal donut fed into the AOS pipeline, with the DL and TIE wavefront estimates compared to the truth from the simulation. The plotted amplitudes are the quantities $s_i\alpha_i$ and $s_i\hat{\alpha}_i$ – i.e., the contribution of each Zernike mode to the PSF FWHM. The error bars on the truth are the expected median 1σ errors due to the atmosphere.

focal donuts, and average the wavefront estimation from each. This is because the TIE solver must operate on intra-/extra-focal pairs. Averaging DL estimates for the same pairs of donuts allows an apples-to-apples comparison of the two algorithms. In the final section, we evaluate the DL model on individual donuts, and discuss the trade-offs with estimation using pairs of donuts.

For all sections below, we define the error as the sum in quadrature of the PSF-weighted wavefront residuals,

$$\text{Error} = \sqrt{\sum_{i=4}^{22} s_i^2(\hat{\alpha}_i - \alpha_i)^2}, \tag{C.5}$$

where $\hat{\alpha}_i$ is the estimated coefficient, and α_i is the truth from the simulation. Due to the

weights s_i, this error is approximately the degradation of the PSF FWHM in arcseconds due to the misestimation of the wavefront (see Appendix C.6).

We note that it takes about 70 ms for the AOS pipeline to return a wavefront estimate when using our DL model, while it takes more than three seconds when using the TIE solver. Thus the DL algorithm is about 40 times faster than the baseline algorithm when estimating the wavefront from a single pair of donuts. Even more substantial speed-ups are possible if you do not constrain the neural network to operate on pairs one-at-a-time like the TIE solver does, as neural networks are very efficient at massive, parallel evaluation. In other words, the DL model enables vectorized wavefront estimation.[2] This speed comes at the up-front computational cost of simulations and network training, but given the goal is to design an algorithm that is able to operate quickly during real-time operations, this trade-off is worthwhile.

C.5.1 Performance on Batoid test set

In this section, we evaluate the performance of the DL and baseline algorithms on pairs of donuts from the Batoid test set, first under ideal conditions, then with varying SNR, vignetting, and blending.

Performance in ideal conditions & comparison to expected atmospheric error

First, we evaluate the performance on an idealized sample that contains no camera vignetting and no bright blends. An example is shown in Fig. C.3. An intra- and extra-focal donut are fed into the AOS pipeline, and both the DL and TIE algorithms are used to estimate the Zernike coefficients. The estimates from each method are shown, along with the true wavefront from the simulations. The error bars on the true wavefront are the median errors expected from the atmosphere (see Appendix C.6). In ideal conditions, an optimal estimator should be consistent with the truth, within the uncertainty from the atmosphere. In this

[2]Note you can parallelize the TIE solver by sending each donut pair to a different processor, but then the number of donuts you can process is limited by the number of processors available for parallel computation.

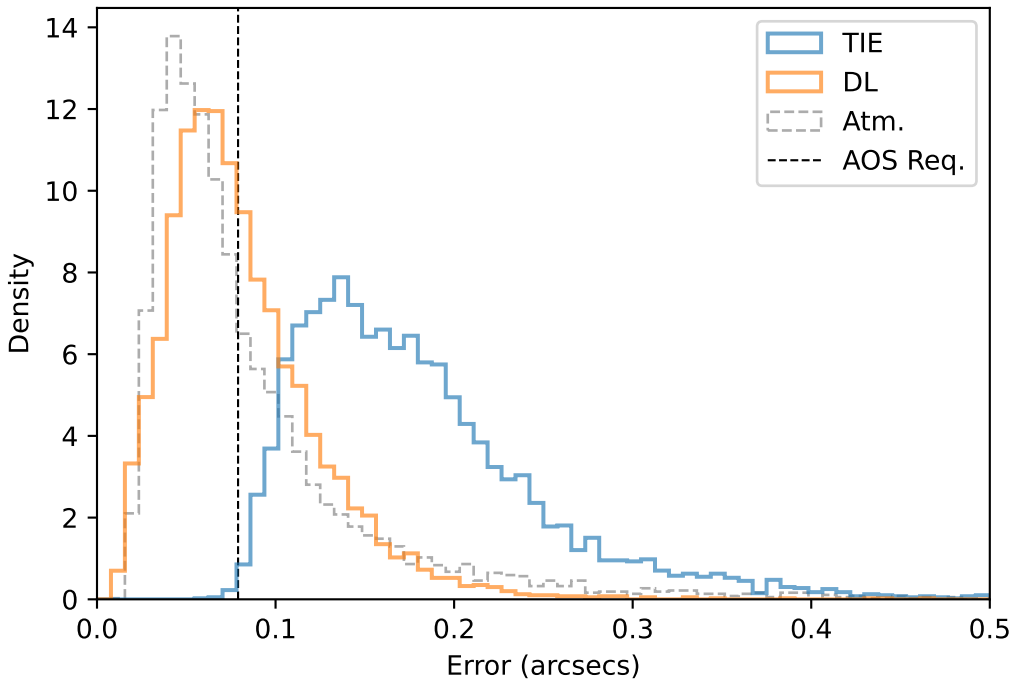

Figure C.4: Distribution of wavefront estimation errors for the DL and TIE methods in ideal conditions (no camera vignetting and no bright blends). The distribution in gray is the expected distribution of wavefront estimation errors due to the phase variance of the atmosphere. The vertical black line is the AOS wavefront estimation error requirement of 0.079".

example, the DL estimate is consistent with the truth within this uncertainty (p-value $\sim$ 0.89), while the TIE estimate is not (p-value ~ 0).

To get a sense of how these methods perform on average, we plot the error distributions for each in Fig. C.4. The DL model systematically outperforms the TIE: the TIE median error is 2 times greater than the DL median, and the low-error tail of the TIE is still on the high-error side of the DL distribution. The expected error distribution from the atmosphere alone is plotted in gray. This distribution is very close to the DL error distribution, suggesting the

DL errors are dominated by the irreducible atmospheric errors. The TIE error distribution, on the the other hand, is shifted to significantly greater errors, indicating a model bias. This is not surprising, as our TIE solver violates several assumptions of the basic TIE method (see Section C.4.1).

The AOS wavefront estimation error requirement [274, 294] of 0.079" is also marked by a vertical black line in Fig. C.4. The DL model is able to meet this requirement 57% of the time using only a single pair of donuts. In contrast, the TIE solver virtually never meets this requirement using only a single pair. Note that you still need a wavefront solution from at least three of the CWFSs in order to interpolate the wavefront solution across the focal plane. This carries the added benefit of averaging over the atmospheric errors that are correlated across any individual CWFS.

Varying SNR

To assess the impact of the image signal-to-noise ratio (SNR), we simulate pairs of stars (isolated, with no camera vignetting) imaged in the Rubin r band with SNR ranging from 10 to 600. We define the total SNR as the total source flux divided by the noise, where the noise includes both the Poisson uncertainty in the source flux and the sky background:

$$\text{noise}^2 = \text{flux} + N_{\text{pixels}} \cdot \sigma_{\text{sky}}^2, \tag{C.6}$$

where $N_{\text{pixels}} = \pi R_d^2(1-\epsilon^2)$ is the number of pixels in a donut. For Rubin, $R_d \approx 66$ pixels is approximately the radius of a donut, and $\epsilon = 0.61$ is the fractional obscuration of the primary mirror.

Fig. C.5 shows the median wavefront estimation error as a function of the SNR, along with the current minimum SNR specified by the Rubin AOS pipeline. The top axis shows the corresponding source magnitude for the median r band sky brightness of 21.20 AB mag arcsec^{-2} [146]. Both methods plateau at high SNR, corresponding to the medians of the distributions in Fig. C.4. Both methods begin to degrade once the total SNR drops below 200. The DL model degrades more slowly, however, and its performance near the AOS SNR limit is similar

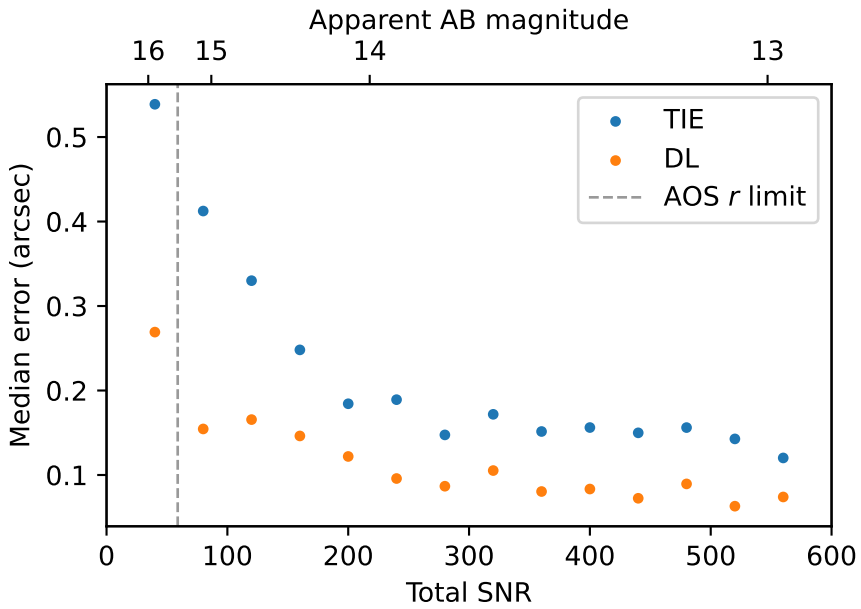

Figure C.5: Median wavefront estimation error as a function of the donut SNR in the r band. The x-axis below the plot shows the total SNR, while the x-axis above the plot shows corresponding r band source magnitude, assuming the median r band sky brightness of 21.20 AB mag arcsec^{-2}. The vertical gray line marks the current minimum SNR specified by the Rubin AOS pipeline.

to the performance of the TIE in the high-SNR regime.

Camera vignetting

In this section, we investigate the impact of camera vignetting on wavefront estimation. All of our donuts experience some amount of vignetting. For donuts at a field angle less than about 1.74°, this is a soft vignetting by the mirror edges. For donuts at a greater field angle, there is sharp vignetting from the body of the Rubin camera. This sharp vignetting impacts 43% of the intra-focal CWFSs, which are farther from the center of the focal plane than the extra-focal CWFSs.

In Fig. C.6, you can see the wavefront estimation errors as a function the field angle of

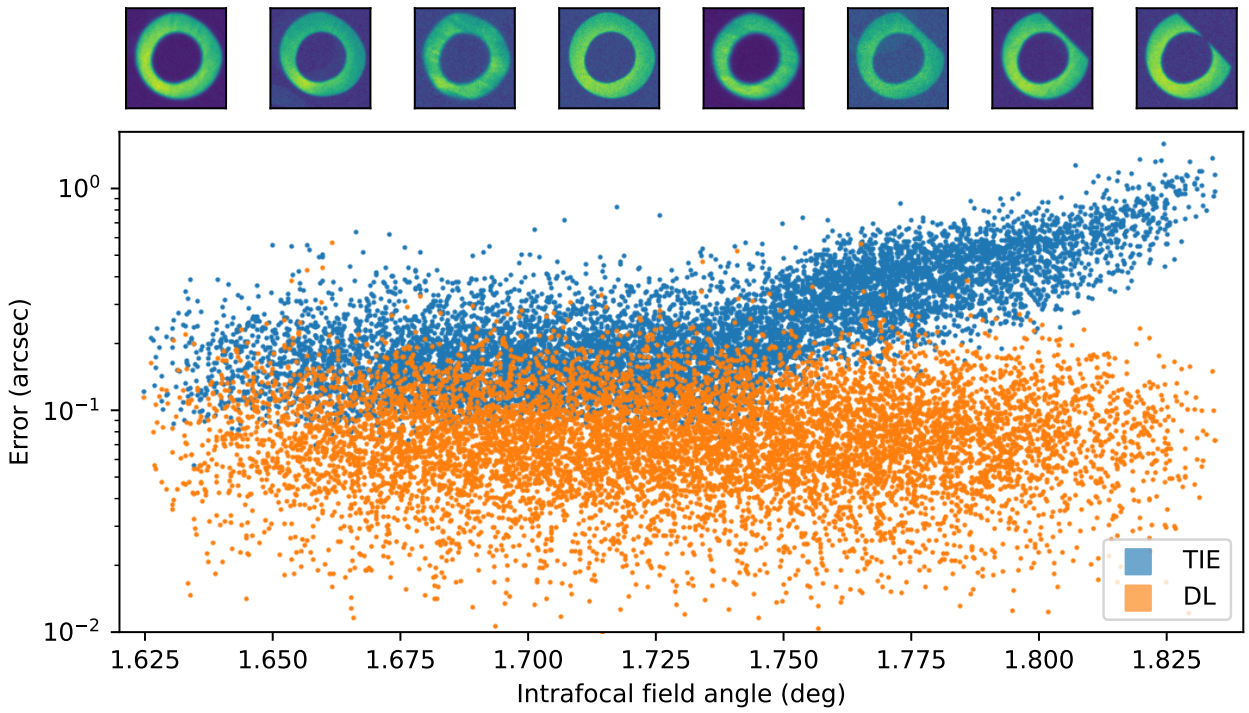

Figure C.6: Wavefront estimation error as a function of intra-focal field angle. Above the scatter plot are donut stamps representative of the given field angle. Sharp vignetting from the camera body impacts field angles greater than about 1.74°. The data set plotted excludes donuts with bright blends.

the intra-focal donut. Above the scatter plot, you can see donut stamps characteristic of the given field angle. Below a field angle of 1.74°, there is soft vignetting from the mirror edges, and beyond this point there is sharp vignetting from the camera body. This transition is associated with a dramatic decline in the accuracy of wavefront estimation with the TIE solver. On the other hand, the DL model is not appreciably impacted by camera vignetting.

This effect can also be seen in Fig C.7, which shows the wavefront error distributions for these two regimes. For the TIE, the camera-vignetted distribution is shifted towards higher errors, while for the DL model, the two distributions are indistinguishable. For camera vignettes, the TIE median error is 5 times worse than the DL median.

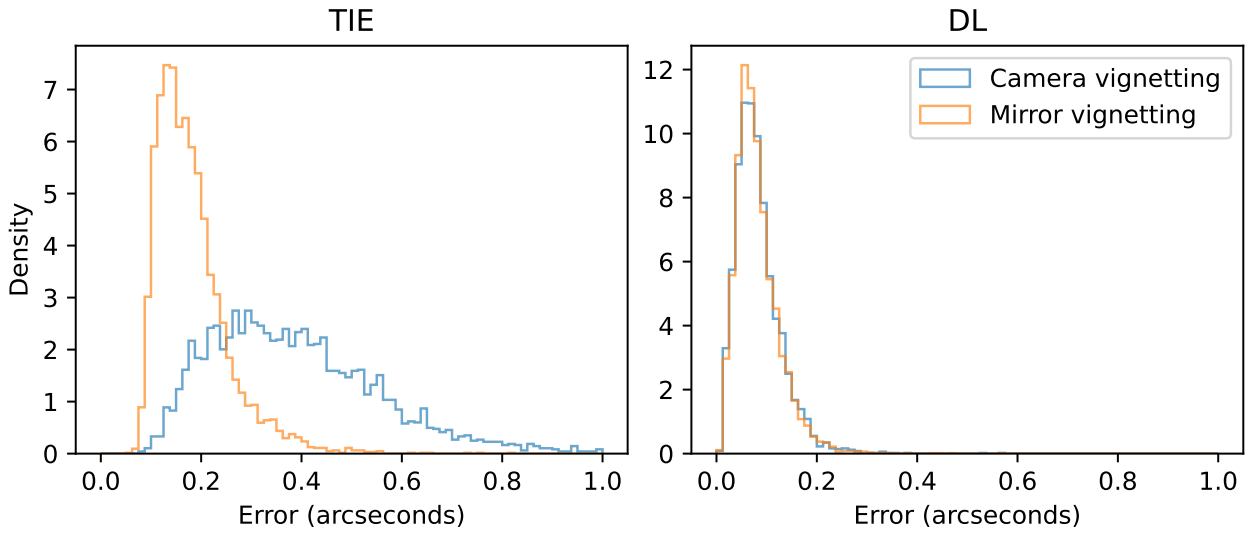

Figure C.7: Left: distribution of errors for the TIE solver, in the presence of mirror and camera vignetting. Right: same for the DL model. The data set plotted excludes donuts with bright blends.

Blending

In this section, we evaluate the impact of blending on wavefront estimation. First, we investigate bright blends, which we define as blends where the overlapping star is within 2 magnitudes of the central star.

Considering only donuts with a single bright blend, in Fig. C.8 we plot the wavefront estimation error vs the distance between the centroids of the central donut and the blending donut. It is evident that the performance of the TIE solver degrades substantially once the blending donut is within approximately 88 pixels of the central donut (for reference, the donut radius is approximately 62 pixels). The left panel of Fig. C.8 shows a blended donut for which the TIE estimate is substantially degraded, with the corresponding TIE error marked with a red cross on the right. You can see the distance at which TIE performance substantially degrades roughly corresponds to the distance at which the central obscurations of the two donuts begin to overlap. Whether the central obscurations for any specific pair of donuts

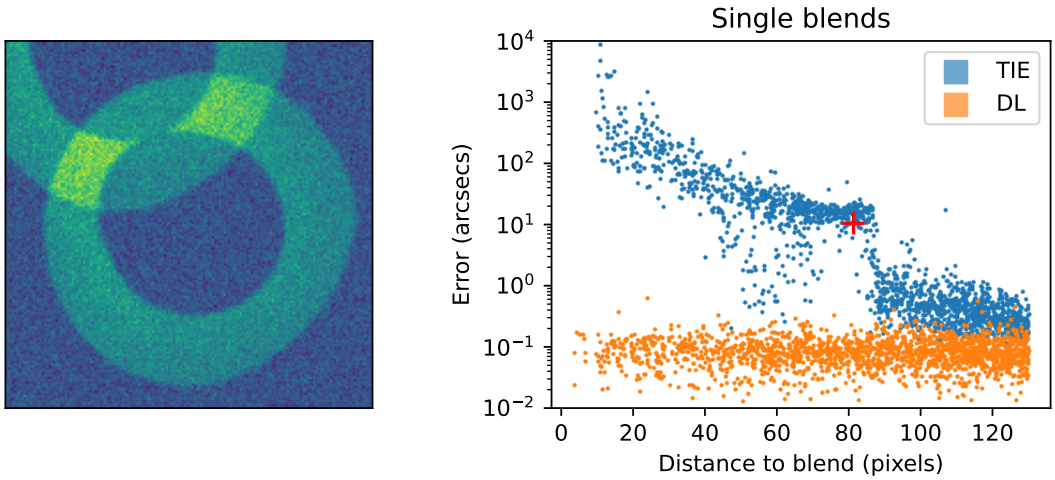

Figure C.8: Left: a bright blend where the blending donut is at a distance of 82 pixels. Right: wavefront estimation error for donuts with a single bright blend, as a function of the centroid-to-centroid distance of the central and blending donut. The red cross denotes the distance and TIE error corresponding to the blend depicted on the left. The data set plotted excludes donuts vignetted by the camera.

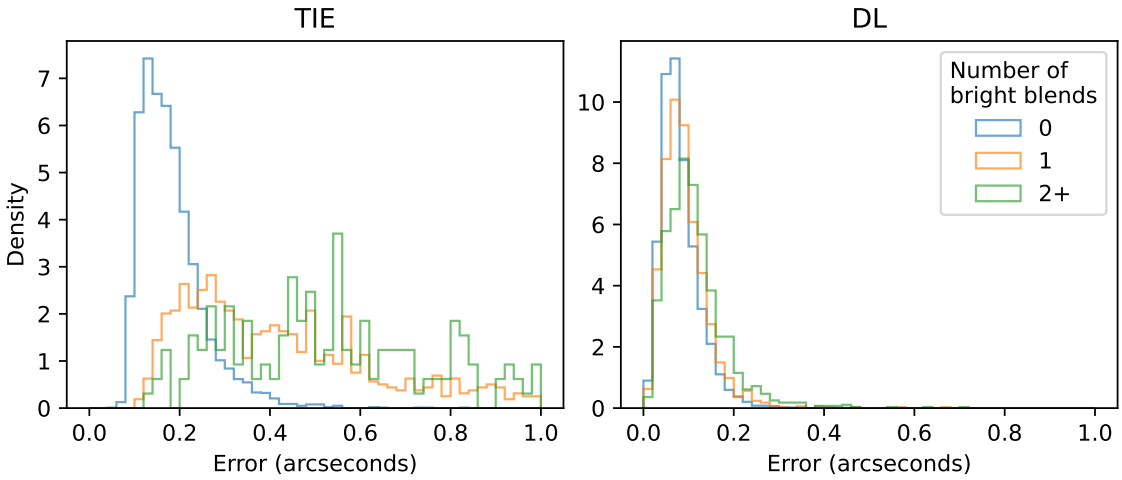

Figure C.9: Left: distribution of errors for the TIE solver for different numbers of bright blends. Right: same for the DL model. The data set plotted excludes donuts vignetted by the camera, and blends for which the centroid-to-centroid distance is less than 88 pixels.

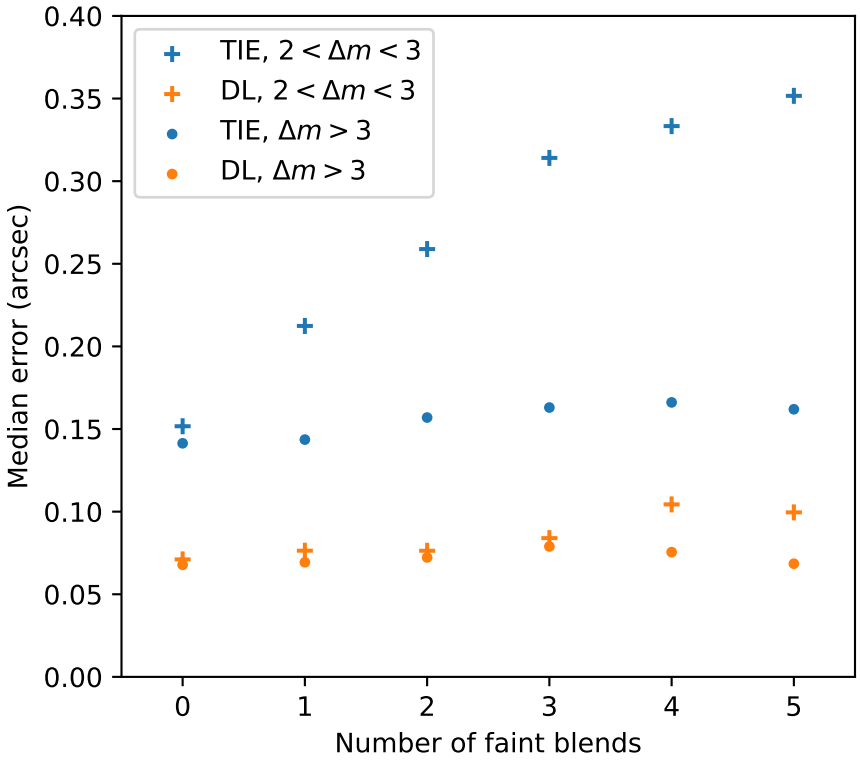

Figure C.10: Median wavefront estimation error as a function of the number of faint blends. The faint blends are divided into two groups: those where the faint blends are at most 1 magnitude beyond the bright blend cut ($2 < \Delta m < 3$), and those where the faint blends are more than 1 magnitude beyond the bright blend cut ($\Delta m > 3$). The data set plotted excludes donuts vignetted by the camera, and donuts with bright blends.

actually overlap at this distance, however, is a function of field angle, orientation of the blend, and the optical perturbations. Interestingly, whether or not the central obscurations actually overlap is not a deterministic predictor of whether the TIE catastrophically fails. Regardless, the DL model appears robust to this transition, with similar performance in both regimes.

After removing donuts with bright blends at a distance less than 88 pixels, we compare the performance as a function of the number of bright blends. In Fig. C.9, you can see

the TIE solver is significantly worse with a single bright blend. With 2 or more bright blends, the TIE solver typically fails catastrophically, because the TIE solver masks blended pixels, and with more than a single blend, too much information is lost. This results in the noisy distribution in Fig. C.9, which has far fewer samples than the unblended and single blend cases. For the DL model, on the other hand, the single blend distribution overlaps the unblended distribution, and the 2+ blend distribution is only slightly degraded. This demonstrates the robustness of the DL model to blending. For single blends where the TIE is relatively successful, the TIE median error is still 14 times worse than the DL median.

Previously, we have only considered bright blends, which were defined as blends within 2 magnitudes of the central star. However, fainter blends essentially modulate the background of the donut, and can also impact wavefront estimation. We define Δm to be the difference between the magnitude of the blending star and the central star, and divide the faint blends into two groups: those where the faint blends are at most 1 magnitude beyond the bright blend cut ($2 < \Delta m < 3$), and those where the faint blends are more than 1 magnitude beyond the bright blend cut ($\Delta m > 3$).

Fig. C.10 plots the median wavefront estimation error as a function of the number of both kinds of faint blends. Both methods are robust to the presence of faint blends that are greater than 3 magnitudes fainter than the central star. However, the TIE method degrades as you add blends between 2 and 3 magnitudes fainter than the central star (these are blends less than 1 magnitude past the bright blend cut). The DL model also experiences some degradation, although to a lesser extent and with more faint blends.

These results indicate the DL model has the potential to dramatically improve AOS performance in very crowded fields where significant blending is unavoidable. These results also provide information that can be used to improve the donut selection criteria for the TIE algorithm.

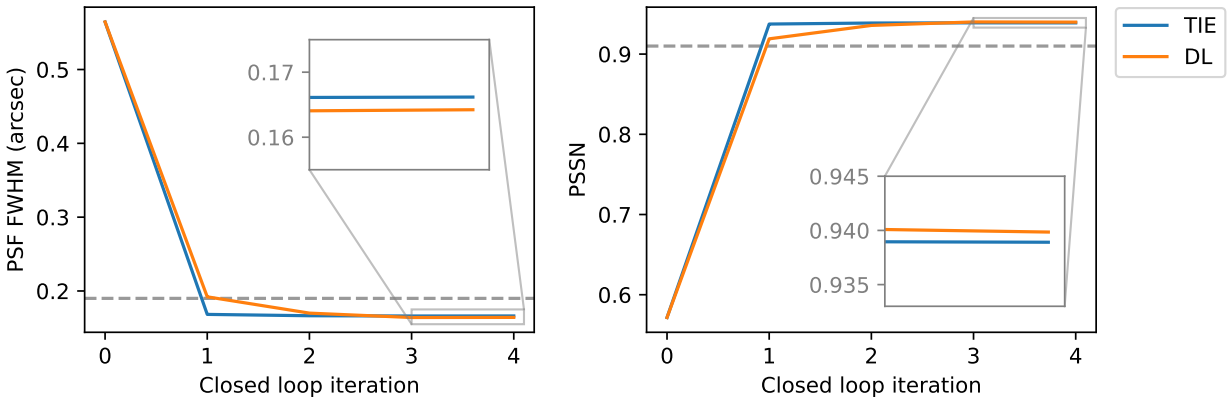

Figure C.11: The PSF FWHM and PSSN as a function of closed loop iteration, when using the DL and TIE methods for wavefront estimation. The AOS requirements are marked by the horizontal dashed lines. The PSF FWHM must be below this limit, and the PSSN must be above this limit. For both metrics, there is an inset that zooms into the last two iterations, where you can see the DL method slightly outperforms the TIE.

C.5.2 Performance on PhoSim closed loop

We wish to understand how the DL wavefront estimator performs when integrated into the full AOS pipeline, including the optical controller that takes the wavefront estimate as an input, and outputs commands to the actuators and hexapods. We also wish to understand how the method performs on data sets not generated with Batoid, to ensure the DL model has not overfit on details specific to Batoid that do not translate to real data. As we do not yet have real images from Rubin, we will perform this test on PhoSim simulations (see Section C.3.2 for details).

We use PhoSim to simulate the AOS closed loop. That is, we start with a simulated telescope with random optical perturbations, and simulate unblended, unvignetted, high-SNR donut images on the intra- and extra-focal chips of each CWFS. We cut out a 160x160 postage stamp of each donut, and use the DL algorithm to estimate the wavefront for each. This information is fed into the optical feedback controller, which uses the wavefronts to

derive optimal forces to apply to each actuator and hexapod, and these commands are applied to the simulated telescope. The updated telescope model is then used to re-simulate images of the same stars, and this process is repeated in a loop. If the estimated wavefront is close to the true wavefront, the optical quality will improve, and on the next iteration, the true wavefront will have smaller perturbations. The estimate of the wavefront perturbations is expected to decrease with each iteration, until the simulated telescope is near perfect alignment and mirror figure. It is important to test whether our DL estimator is capable of estimating wavefronts that drive the system towards and maintains the required optical quality.

The two metrics we track here are the PSF FWHM and the PSSN (Normalized Point Source Sensitivity). The PSF FWHM represents the resolution of imaging, and PSSN is related to the relative SNR of observations and therefore represents the efficiency loss in telescope observing time [276]. The AOS closed loop is required to converge to a FWHM of less than 0.19" and a PSSN greater than 0.91 (see Appendix C.6).

The PSF FWHM and PSSN as a function of closed loop iteration are plotted in Fig. C.11 for the DL and TIE algorithms. In these simulations, both algorithms meet the requirements, which are plotted in gray. Furthermore, after reaching their optimal values, both algorithms maintain the same optical quality on subsequent iterations, demonstrating the robustness of the DL model to atmospheric noise and the stability of the system when working together with the optical controller. In the inset zoom-in's, you can see the DL model slightly outperforms the TIE on both metrics. This improvement is modest, however, indicating the real strength of the DL method will not be in the case where you have an unblended, unvignetted, high-SNR donut on every wavefront sensor, but rather in the cases where one or multiple of these optimistic assumptions are broken. Testing closed loop convergence in less optimal conditions is a future goal of the Rubin active optics team.

This test also demonstrates the robustness of the DL model to the differences between PhoSim and Batoid that are described in Section C.3.2, and provides a check that training the DL model to optimize the PSF FWHM does not come at the expense of the PSSN (there

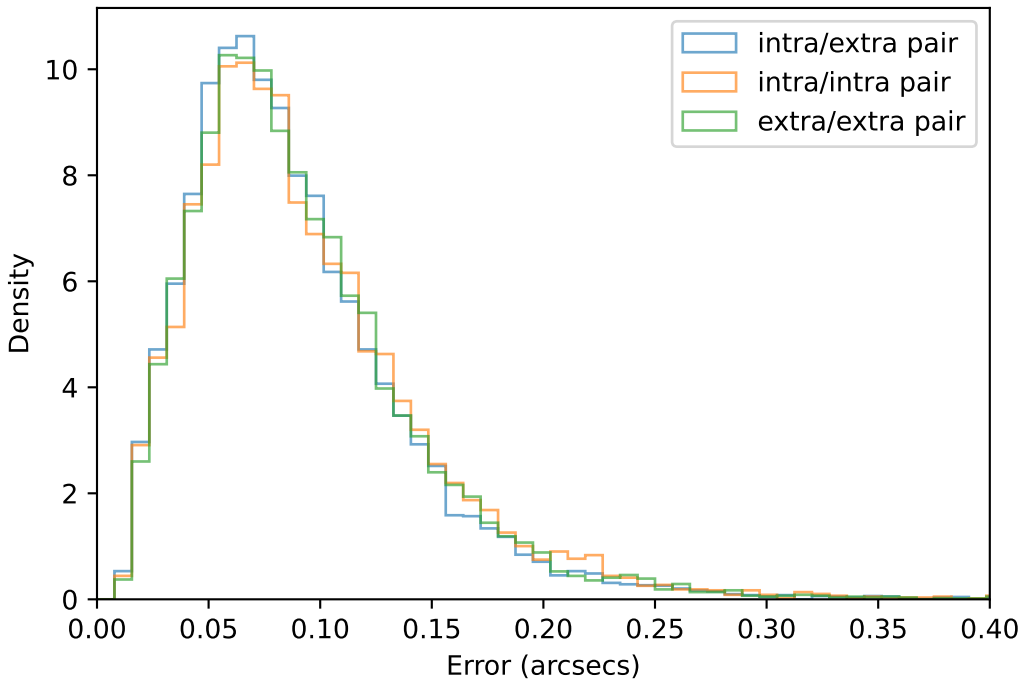

Figure C.12: Distribution of wavefront estimation errors for the DL model, using intra/extra, intra/intra, and extra/extra pairs. These distributions coincide, indicating that the DL model can accurately estimate the wavefront only using information from one side of focus.

is more discussion of this in Appendix C.6). However, while both algorithms converge to almost the same optical quality, the DL algorithm requires an extra iteration to converge. The implications of this extra step are discussed in Section C.6.

C.5.3 Estimating the wavefront using a single donut

In the preceding sections, we used the DL model to estimate the wavefront using a pair of intra- and extra-focal donuts, enabling an apples-to-apples comparison with the TIE solver. The DL model, however, can estimate the wavefront using only a single donut, meaning it does not need information from both sides of focus. To evaluate DL performance on one side

of focus, we calculate wavefront estimation errors while averaging the wavefront from pairs of donuts *on the same side of focus*. We do this pairing to make sure we are still getting the benefit of noise reduction from averaging a pair of wavefronts, while only using information from one side of focus. You can see in Fig. C.12 the DL model does well with *only* intra-focal or extra-focal images. Interestingly, it appears that averaging two wavefronts only reduces the noise by a factor of ~ 1.2, 15% less than the factor of $\sqrt{2} \sim 1.41$ expected for uncorrelated Gaussian error. This is due to correlation of atmospheric errors across individual CWFSs due to low-altitude turbulence.

The ability to estimate the wavefront using only images from one side of focus has several advantages. For example, in crowded fields, you may not be able to find an intra-/extra-focal pair where each donut is sufficiently bright, and is not too heavily blended or vignetted. It removes the need to compensate for differences between the two sources in a pair, which the baseline algorithm must do in order to apply the TIE. It can also provide more localized wavefront estimates across each CWFS, which provides more leverage when interpolating the wavefront across the full focal plane. We note that the methods of A. Tokovinin & S. Heathcote [299], R. J. Janish [152], and A. Roodman et al. [244] are all also capable of estimating the wavefront using an image from only one side of focus.

There may also be downsides, however, to using information from a single side of focus. Distortion of the pupil, attenuation of the mirror, and differential chromatic refraction (DCR) all have effects that are independent of the side of focus on which a donut is imaged. Wavefront aberrations, on the other hand, change parity as they pass through focus. For example, astigmatism and DCR both elongate sources, but the former rotates by 90°when you pass through focus while the latter is always oriented towards zenith. Using information from both sides of focus allows you to distinguish these effects. We tested the impact of DCR by simulating very blue stars at airmass 4, imaged in the Rubin g band, while toggling the simulated DCR on and off. We find that in this extreme case, DCR increases the median wavefront estimation error by 7% when using information from only one side of focus, while the median only increases by 5% when averaging estimates from both sides of focus. This

is only a modest increase, even in a situation where DCR will be the most severe. Furthermore, if this does become a limiting factor in certain circumstances, the effects of DCR can be predicted and separately modeled.

C.6 Discussion & Conclusion

In this chapter, we have built a deep learning (DL) model for wavefront estimation. Previous works have built prototypes for this application [292, 293, 319], but this is the first time such a system has been integrated into the Rubin AOS pipeline, and validated against the baseline algorithm. We have demonstrated that under ideal conditions, where the images do not have bright blends and are not vignetted by the camera, the DL system outperforms the baseline model, improving the median wavefront estimation error by a factor of 2. In fact, the estimation error for the DL model is comparable to the expected minimum error due to random phase fluctuations from the atmosphere. This indicates that the DL system is performing as good as can be expected when given high-quality data.

Furthermore, the DL system is far more robust than the baseline algorithm when presented with less ideal data. When the intra-focal donut is too far from the center of the focal plane, its edge is sharply vignetted by the camera body. This results in an exponential decrease in accuracy for the baseline algorithm, while the DL model is unaffected. Furthermore, blending of the central source with neighboring sources greatly diminishes the accuracy of the baseline method, and often causes catastrophic failures. Again, the DL model is largely unimpacted, improving the median wavefront estimation error by a factor of 14. Both algorithms begin to degrade when the total SNR of the source drops below 200, but the degradation of the DL algorithm is less severe.

In addition to the strengths listed above, the DL model can accurately estimate the wavefront using only a single image, whereas the baseline algorithm requires a pair of intra- and extra-focal images. The DL model is also approximately 40 times faster when, like the baseline algorithm, it is applied to pairs one-at-a-time, but much greater speed-ups are achievable if the DL model is applied to many images simultaneously.

These strengths make the DL model a valuable addition to the Rubin AOS. It will allow the AOS closed loop to operate in the densest 8% of fields where the baseline algorithm fails due to the inability to select a pair of unvignetted and unblended sources. This will expand by 1400 $\deg^2$ the area useful for precision galactic science that targets the crowded plane of the Milky Way. Higher performance with degraded data and the ability to estimate the wavefront given a single source from each CWFS also increases the robustness of the AOS to a wide variety of conditions, such as satellite trails and failing CCDs.

The one test where the baseline algorithm outperforms the DL model is the convergence rate of the AOS closed loop simulated with PhoSim. PhoSim is a different raytracer than that used to simulate the training data for the DL model. The DL model still meets the closed loop requirements and converges to the same optical quality, but requires an additional closed loop iteration to do so. This might be due to chance, and simulating many more closed loops in a wide variety of conditions may show that the DL model exceeds the baseline on this test as well. However, these results might also indicate a need to simulate a wider range of wavefront perturbations, as the initial perturbation in the closed loop might lie outside the area of parameter space covered by our training set.

To prepare for commissioning, we plan to generate a larger library of simulations. These simulations will more densely and expansively sample the space of potential optical perturbations, as well as include a larger number of independently simulated atmospheres. This will reduce the dangers of overfitting and network extrapolation. We also plan to include more realistic estimates for stellar temperatures in order to test for chromatic effects, and to upgrade our simulations to include more image effects, such as mirror print through, diffraction from the telescope support structure, cosmic rays, and bad pixels.

We expect that domain adaptation, the technique of adapting a model trained on one data set so that it performs well on a different target data set, will also be necessary for commissioning our deep learning system. This is because real data is always messy in ways not captured by simulations, and deep learning systems need to develop robustness to these differences in order to perform accurate inference on real data. This task will be difficult for

the Rubin AOS, since the vast majority of Rubin images will be "unlabeled" in the sense that the true wavefront will not be known[3]. For this reason, we expect strategies such as domain adversarial training [54, 55] will be valuable when commissioning this system on real data. Domain adversarial training uses a discriminator network that takes the CNN image features as input, and tries to guess whether the original image was real or simulated. By training the CNN to outsmart the discriminator, the CNN learns to ignore artifacts and features that are unique to the real images, and instead focus on the physical information present in both the real images and the simulated training set.

It may also be advantageous to further optimize the network architecture for domain adaptation, since two of our architectures achieved similar performance on the Batoid test set, but differing performance in the PhoSim closed loop. Overcoming the obstacle of domain adaptation will be the greatest challenge during commissioning. We have begun working towards these goals using data from the Rubin Observatory Auxiliary Telescope (AuxTel), which is a simpler system and is already producing data.

Finally, the results presented in Section C.5.1 will also be useful in improving the baseline wavefront estimation algorithm. Using the wavefront estimation error as a function of SNR, vignetting, blend distance, and the number of faint blends will help refine the donut selection criteria. The evidence of a model bias compared to the atmospheric uncertainty and the DL performance motivates additional investigation into the modeling assumptions of the baseline algorithm.

Appendix C.a: Clarifying the LSST Optical Error Budget

Throughout this chapter, there are several references to the LSST error budget. The numbers cited are closely related, but subtly different. In this section we clarify the relationship between these numbers. Note that when combining errors from multiple sources, PSF FWHM

[3]The exception to this will be a small number of exposures for which we will apply large, known perturbations to the telescope control parameters, so that the applied perturbations will dwarf any natural, unknown perturbations.

values are meant to be added in quadrature and PSSN values are meant to be multiplied.

In Section C.2, we state:

> The median seeing at Rubin's site on Cerro Panchón is 0.65" [146], and Rubin's optical system is required to degrade this value by no more than 0.4".

This 0.4" requirement, from C. F. Claver & LSST Systems Engineering Integrated Project Team [57], is for the entire optical system, including all contributions from both the telescope and camera. In particular, of this 0.4" requirement, 0.25" is allocated to the telescope, 0.3" is allocated to the camera, and 0.08" is allocated to the optical design.

Later in this same section, we state:

> To deliver the optical quality required for LSST, the AOS must limit the PSF (Point Spread Function) FWHM (Full Width Half Maximum) contribution of optical aberrations to no more than 0.09", including only 0.079" due to misestimation of the wavefront – less than 10% of the total error budget.

The 0.09" requirement is the portion of the telescope error budget allocated to the active optics system, and of this, 0.079" is allocated to optical errors due to misestimation of the wavefront. These numbers are sourced from S. Thomas et al. [294] and J. Sebag & LSST Systems Engineering Integrated Project Team [274].

In Section C.5.2, we state:

> The AOS closed loop is required to converge to a FWHM of less than 0.19" and a PSSN greater than 0.91.

The 0.19" requirement cited here contains the 0.09" allocated to the active optics, plus additional errors of 0.167" allocated to M1M3 and 0.032" allocated to M2 (which are also captured by the closed loop simulation). The quoted PSSN (0.91) is similarly the product of the PSSN specified for the active optics (0.981), M1M3 (0.933), and M2 (0.998). These numbers are also sourced from J. Sebag & LSST Systems Engineering Integrated Project Team [274].

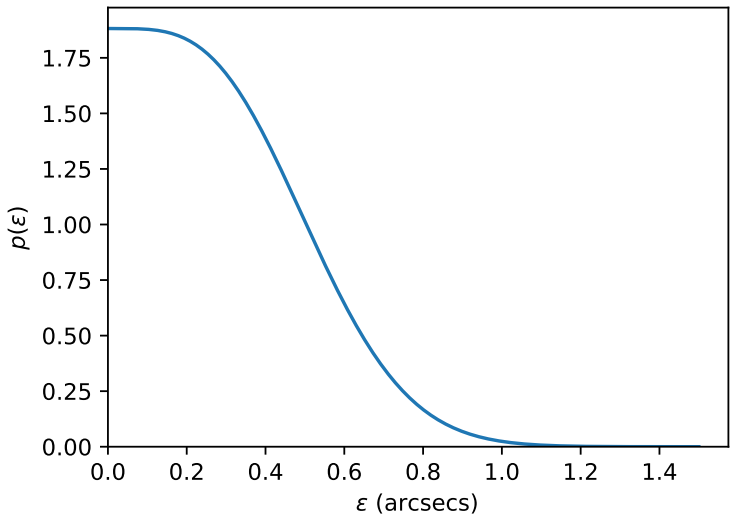

Figure C.13: The distribution of PSF FWHM degradations sampled for the simulations.

Appendix C.b: Sampling Rubin optical perturbations

For each simulated telescope pointing, we randomly perturb the 50 parameters that specify the optical alignment and the figure of the mirror. Parameters 1-5 and 6-10 describe the rigid body motions of M2 and the camera, respectively. These include translations along all three axes (in microns), and rotations about the x and y axes (in arcseconds). Parameters 11-30 and 31-50 are the bending mode amplitudes in microns for M1M3 and M2, respectively.

To perturb these parameters, we first sample a PSF FWHM degradation from a folded Gaussian distribution whose standard deviation is equal to its mean (i.e., $\sigma = \mu$). This distribution was heuristically chosen to yield a distribution that emphasizes small degradations, but has a tail towards much larger degradations. The PDF of this distribution is

$$p(\varepsilon) = \frac{1}{\sqrt{2\pi\mu^2}} \left(e^{-(\varepsilon/\mu - 1)^2/2} + e^{-(\varepsilon/\mu + 1)^2/2} \right), \tag{C.7}$$

where ε is the PSF FWHM degradation in arcseconds. This distribution is plotted in

Fig. C.13. The parameter μ is chosen such that $\langle\varepsilon\rangle = 0.3$", using the formula

$$\mu = \langle\varepsilon\rangle \left[\sqrt{\frac{2}{\pi}} e^{-1/2} + \text{erf}\left(\frac{1}{\sqrt{2}}\right)\right]^{-1}. \tag{C.8}$$

In other words, we sample ε from a Gaussian with mean and standard deviation equal to μ, then take the absolute value. This determines the total PSF FWHM degradation associated with the perturbation. We then determine what portion of this PSF degradation to assign to each telescope parameter by sampling from the 50-dimensional hypersphere with radius ε. Finally, we use the values of ΔPSF/Δperturbation determined in Appendix C.6 to convert these PSF degradations to parameter perturbations in the appropriate unit.

We use Batoid to simulate images from telescopes with these simulated perturbations. For each simulation, we use Batoid to extract the wavefront and calculate the corresponding Zernike coefficients. As described in Section C.4, these wavefronts are at the center of each CWFS, for a fiducial wavelength of 1 μm, and have the intrinsic aberrations of the telescope design subtracted.

Appendix C.c: Estimating PSF FWHM degradation

It is useful to express wavefront estimation errors in a unit that is readily interpretable in terms of the optical quality. In this appendix, we derive a set of scaling factors, s_i, such that a given wavefront perturbation $\varphi = \sum_i \alpha_i Z_i$ results in an increase of the PSF FWHM by the amount

$$\text{PSF FWHM increase} = \sqrt{\sum_i (s_i \alpha_i)^2}. \tag{C.9}$$

To derive these factors, recall that the gradient of the wavefront provides a map from pupil position to focal plane position (e.g., C. Roddier & F. Roddier 242). We can express the gradients of the wavefront with two more Zernike expansions:

$$\frac{\partial\varphi}{\partial x} = \sum_i \alpha_i^x Z_i \qquad \frac{\partial\varphi}{\partial y} = \sum_i \alpha_i^y Z_i, \tag{C.10}$$

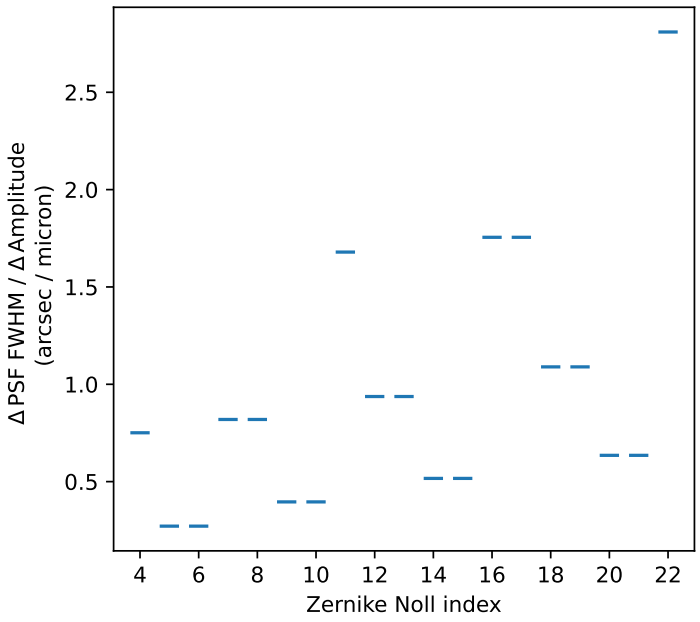

Figure C.14: The PSF FWHM weighting factors, s_i for the Rubin Observatory.

the coefficients of which can be determined analytically from the expansion coefficients of φ [222, 283, 322] or calculated using the Galsim Zernike library [245]. The variance of the x position of photons on the focal plane is then

$$\sigma_x^2 = \frac{1}{A}\int d^2\rho \left[\frac{d\varphi}{dx}(\rho)\right]^2 = \frac{1}{A}\sum_{i,j}\alpha_i^x\alpha_j^x \int d^2\rho\, Z_i(\rho)Z_j(\rho) \tag{C.11}$$

$$= \sum_{i,j}\alpha_i^x\alpha_j^x\,\delta_{ij} = \sum_i(\alpha_i^x)^2, \tag{C.12}$$

where ρ is a 2D position vector on the pupil and we use the Zernike normalization defined in Eq. C.2. The variance in y positions is calculated analogously. Fitting an isotropic Gaussian to the photon positions on the focal plane yields the variance

$$\sigma^2 = \frac{1}{2}(\sigma_x^2+\sigma_y^2) = \frac{1}{2}\sum_i(\alpha_i^x)^2+(\alpha_i^y)^2. \tag{C.13}$$

Finally, the FWHM can be calculated with the regular scaling for a Gaussian:

$$\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\sigma \tag{C.14}$$

Scale factor s_i can be calculated in units of arcseconds per micron by applying the above formulae to the wavefront $\varphi = \alpha_i Z_i$, where $\alpha_i = 1$ μm. The values of s_i for the Rubin Observatory are shown in Fig. C.14. All of the scale factors are of $\mathcal{O}(1)$. Note that the values of these scale factors depend on the geometry of the pupil, and are therefore telescope specific.

We used simulations to validate our Zernike-to-PSF-FWHM conversion scheme. First we simulated a pair of in-focus stars: one with an unaberrated wavefront, and one with random aberrations in Zernikes 4 to 23. The random aberrations were independently drawn from a Gaussian distribution — i.e. the aberrations in each mode have identical variance, but no covariance between different Zernike modes. For each pair of stars, we used Galsim to measure the PSF FWHM, and took the difference to estimate the increase to the FWHM due to the wavefront aberrations. We then calculated the expected FWHM increase from the Zernike amplitudes, using the method described above.

The results of this experiment are shown in Fig. C.15. For FWHM degradations $\lesssim 0.25$ arcsecs, the Zernike conversion scheme described in this appendix is a good match to the degradations measured from image simulations. Above this level, our scheme mildly overestimates the FWHM degradation, likely due to degeneracies (and therefore cancellations) in the way each Zernike mode impacts the PSF FWHM, which are not captured by our method. This is not a significant problem as Fig. C.4 shows the majority of our wavefront errors correspond to PSF FWHM degradations below the 0.25 arcsecond level. It can be assumed that the errors in the high error tail of Fig. C.4 are slightly over estimated. This caveat, however, does not change the conclusions of the main text.

Note that while the majority of Zernike estimation errors lie below the 0.25 arcsecond level, the same is not necessarily true of the Zernike amplitudes themselves. It is still useful to plot Zernike amplitudes in units of PSF FWHM contribution, however, as this enables a by-eye estimate of how the Zernike estimation residuals impacts the PSF of the telescope.

Similarly to the Zernike amplitude conversions described above, you can calculate PSF FWHM degradation for each of the optical parameters discussed in Appendix C.6. We

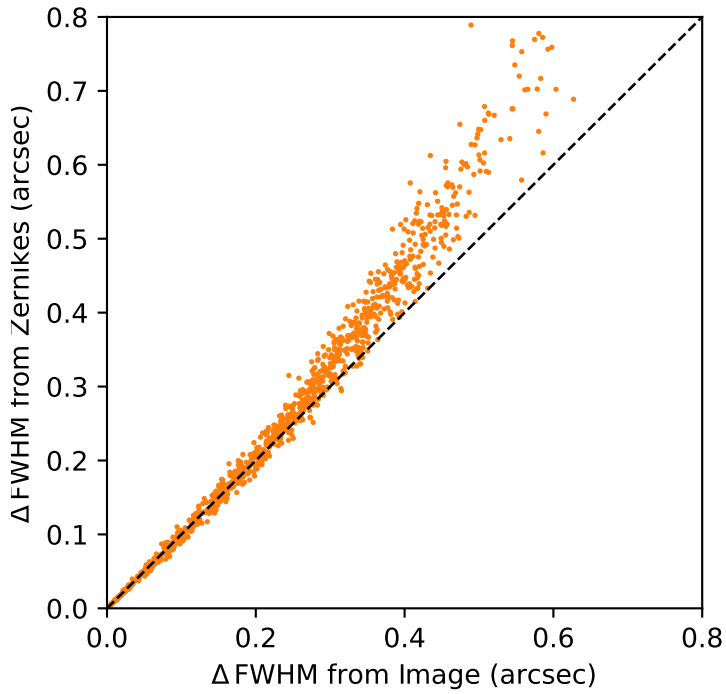

Figure C.15: Comparing increase in PSF FWHM due to wavefront aberrations for 10^3 simulated stars with random Zernike amplitudes. The x-axis shows increase in FWHM as measured from simulated images; the y-axis shows the value estimated from our weighted Zernike scheme; the dashed line indicates perfect correspondence. The two methods agree for FWHM increases below 0.25", but above this level, the weighted Zernike scheme overestimates the increase in FWHM.

perturb each parameter 1 unit, use Batoid raytracing to retrieve the corresponding wavefront, and then proceed as above.

Appendix C.d: Simulating the atmosphere

In addition to distortions from misalignment and mirror deformation, the Rubin optical wavefront is distorted by atmospheric turbulence. Because Rubin's field of view is so large,

we cannot correct for this turbulence. In effect, the atmosphere adds a random Zernike series to the Zernike series we are trying to estimate. These random phase variations from the atmosphere are an irreducible noise floor for wavefront estimation, so it is useful to understand their statistics.

Using Eq. C.2, we can write the covariance of the coefficients α_i as

$$\langle \alpha_i \alpha_j \rangle = \frac{1}{A^2} \int d^2\rho \int d^2\rho' \; Z_i(\rho) Z_j(\rho') \, \langle \varphi(\rho) \varphi(\rho') \rangle. \tag{C.15}$$

We assume the von Kármán turbulence model [306], in which turbulence begins at an upper scale L_0 and cascades down to the viscous dissipation scale ℓ_0. The phase covariance for von Kármán turbulence is [43]

$$\langle \varphi(\rho) \varphi(\rho') \rangle = 0.0858 \left(\frac{L_0}{r_0} \right)^{5/3} \left(\frac{2\pi r}{L_0} \right)^{5/6} K_{5/6} \left(\frac{2\pi r}{L_0} \right), \tag{C.16}$$

where $r \equiv |\rho - \rho'|$, and K is a modified Bessel function of the second kind. r_0 is the usual Fried parameter [108], which, for von Kármán turbulence, is related to L_0 and the PSF FWHM via the formula [298]

$$\mathrm{FWHM} = \frac{0.976\lambda}{r_0} \sqrt{1 - 2.183 \left(\frac{r_0}{L_0} \right)^{0.356}}. \tag{C.17}$$

We assume individual turbulent cells do not evolve during the 15 second exposures, and so the coherence time of turbulence is determined by the wind speed, which moves new turbulent cells across the field of view during an exposure. The coherence time is approximately $\tau_0 = 0.31\, r_0 / v_e$, where v_e is the average wind speed, weighted by the strength of turbulence in each layer [243]. Thus, during a 15 second exposure, the phase covariance is reduced by a factor of $N = 15\, s/\tau_0$.

We use GalSim [245] to simulate an atmosphere with von Kármán turbulence in 6 layers. We use the turbulence profile measured on Cerro Armazones [99] (with 10% Gaussian noise). Outer scales L_0 are randomly drawn from a log-normal distribution with mean 30 m and standard deviation 20 m, truncated between 10 and 100 m. Wind speeds are isotropically

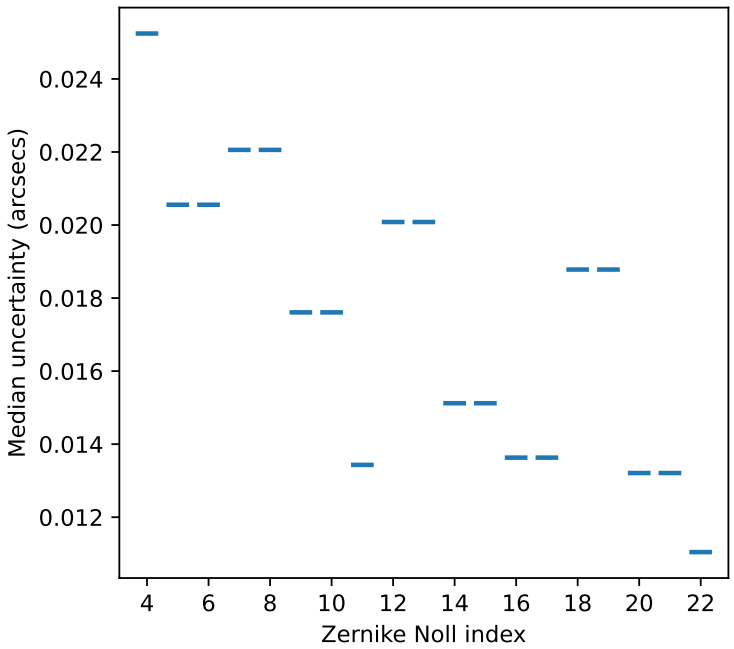

Figure C.16: Median 1σ error due to the atmosphere for each Zernike coefficient α_i.

sampled with a uniform distribution between 0 and 20 m/s. We assume constant air pressure of 69 kPa, H_20 pressure of 1 kPa, and temperature of 293 K.

To estimate the impact of turbulence on Zernike estimation, we numerically evaluate Eq. C.15 while sampling the PSF FWHM and wavelength from the baseline LSST OpSim simulation, and L_0 and v_e from the distributions described above. The median standard deviation of each Zernike coefficient is shown in Fig. C.16. The sum in quadrature of these values is 0.072". For a comparison of the wavefront estimation algorithm errors to the atmospheric error, see Section C.5.1.

R. J. Noll [222] studied the projection of atmospheric statistics onto Zernike coefficients for a circular aperture (i.e., no central obscuration) and Kolmogorov turbulence (the $L_0 \to \infty$ limit of Eq. C.16; D. L. Fried 108). Their study found that for each coefficient, the variance scales like $r_0^{-5/3}$, and that the total variance contained by all coefficients greater than Noll index j scales like $j^{-\sqrt{3}/2}$ for large values of j. Despite studying von Kármán turbulence for

an aperture with a large central obscuration, we find the same scaling relations (for fixed L_0, before dividing by N). We note that the variance averaged over a 15 second exposure scales like $r_0^{-2/3}$, as this entails a division by $N \propto r_0^{-1}$.

Appendix C.e: Additional wavefront estimation metrics

In this appendix, we provide a few more figures to check that the conclusions of this chapter do not depend on the choice of PSF FWHM as the metric for wavefront estimation, as well as a few figures that check the consistency of our results when selecting different subsets of the test set.

Fig. C.17 is a near duplicate of Fig. C.4, except that the wavefront residuals are measured in nanometers. This provides a check on our conversion to PSF FWHM by showing that even without this weighting scheme, the DL method outperforms the TIE. The shapes of the histograms in these two plots are nearly identical, indicating that the PSF FWHM carries nearly the same information as the residual in nanometers, with the benefit of having a unit that is more readily interpretable in terms of optical quality.

Fig. C.18 is another histogram comparing the performance of the DL and TIE methods on pairs of stars (with no camera vignetting or bright blends). This figure shows PSSN values after correcting for simulated aberrations using the estimated wavefront from each method. Recall that PSSN ranges between zero and one, with greater values indicating better optical quality. This figure shows that the DL method also outperforms the TIE method on maximizing the PSSN, providing a check that training to optimize the PSF FWHM does not come at the expense of other optical quality metrics. Interestingly, both the DL and TIE methods achieve the PSSN required by the error budget (marked with the dashed line) using a single pair of donuts more often than they achieve the required PSF FWHM (cf. Fig. C.4).

Fig. C.19 provides a check that both methods have similar performance on each of the wavefront sensors. There is no reason to expect different performance on different sensors, so this plot provides a consistency check on the simulations and the training of the DL method.

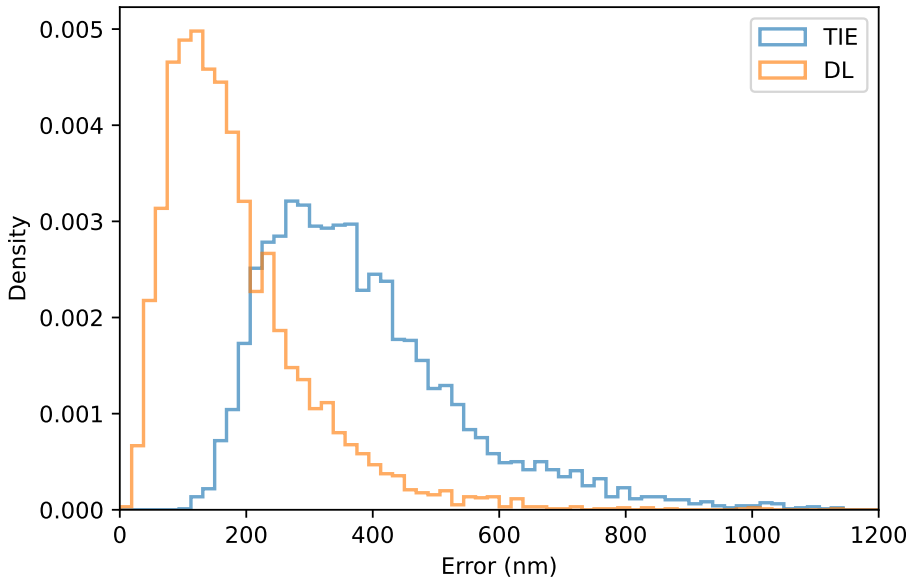

Figure C.17: Distribution of wavefront estimation errors for DL and TIE methods in ideal conditions (no camera vignetting and no bright blends), in units of nanometers. This is a near duplicate of Fig. C.4, except with a different unit on the x-axis.

Fig. C.20 shows the performance for each LSST filter. The errors are comparable between every band, with slightly worse performance for the u, z, and y bands. This is expected as the average SNR in these bands is slightly worse.

Fig. C.21 compares the error distribution for each Zernike coefficient. This shows that the DL method outperforms the TIE method on wavefront estimation for every Zernike mode. In addition, compared to the TIE, the DL method has less variation in performance for different Zernike modes.

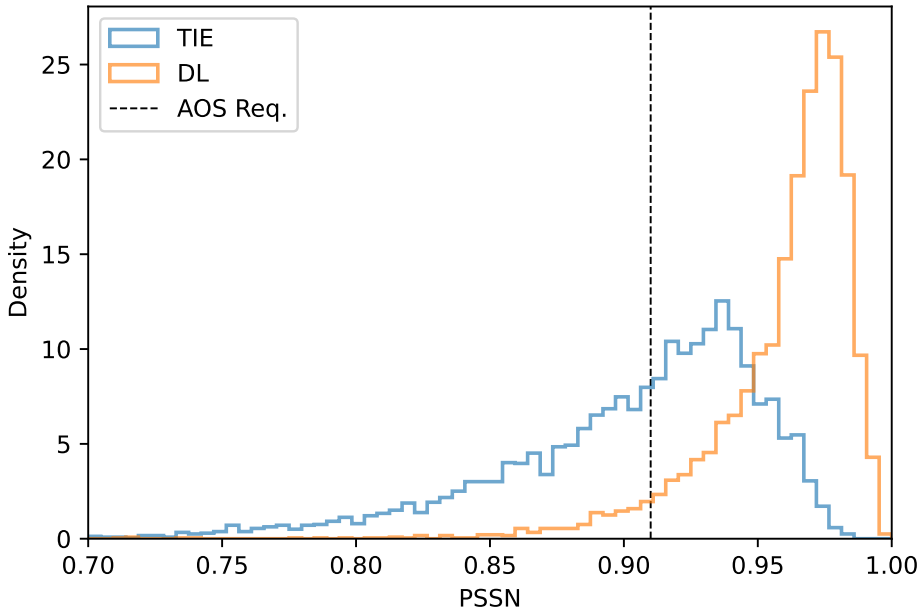

Figure C.18: Distribution of PSSN values after correcting for aberrations using the estimated wavefront from the DL and TIE methods. The vertical black line is the AOS wavefront estimation error requirement of 0.91.

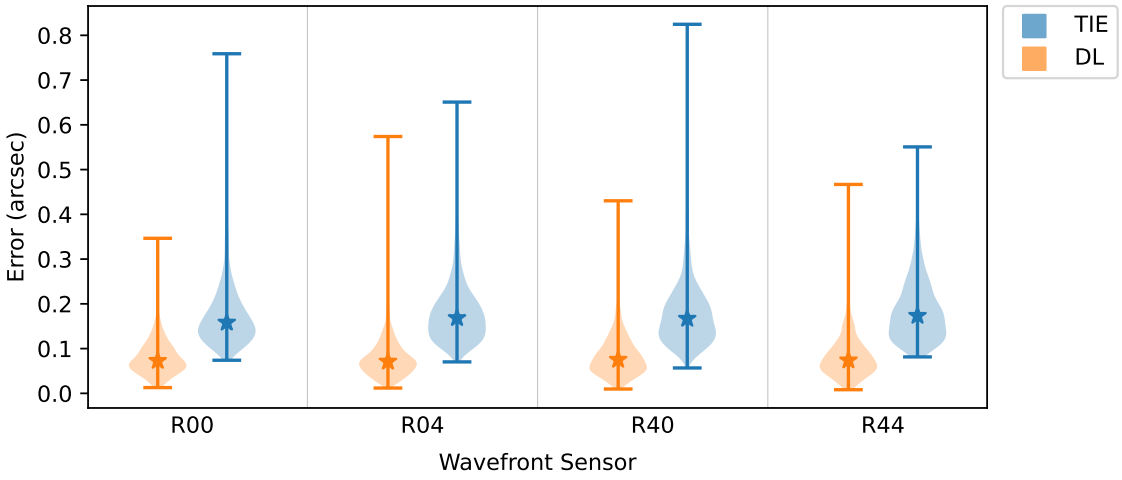

Figure C.19: Distribution of wavefront estimation errors for each CWFS, labeled according to the Rubin convention. The stars denote the median error.

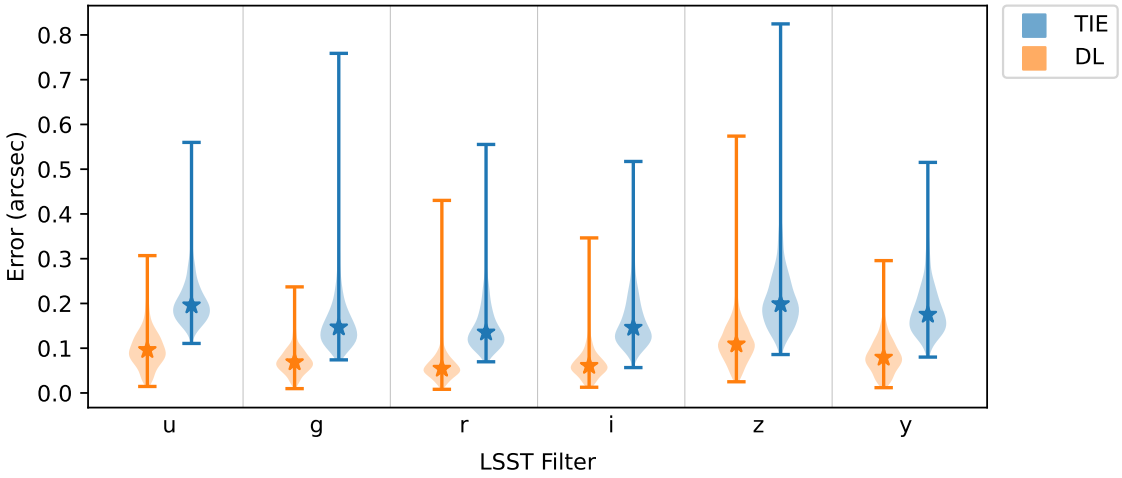

Figure C.20: Distribution of wavefront estimation errors for each of the LSST filters. The stars denote the median error.

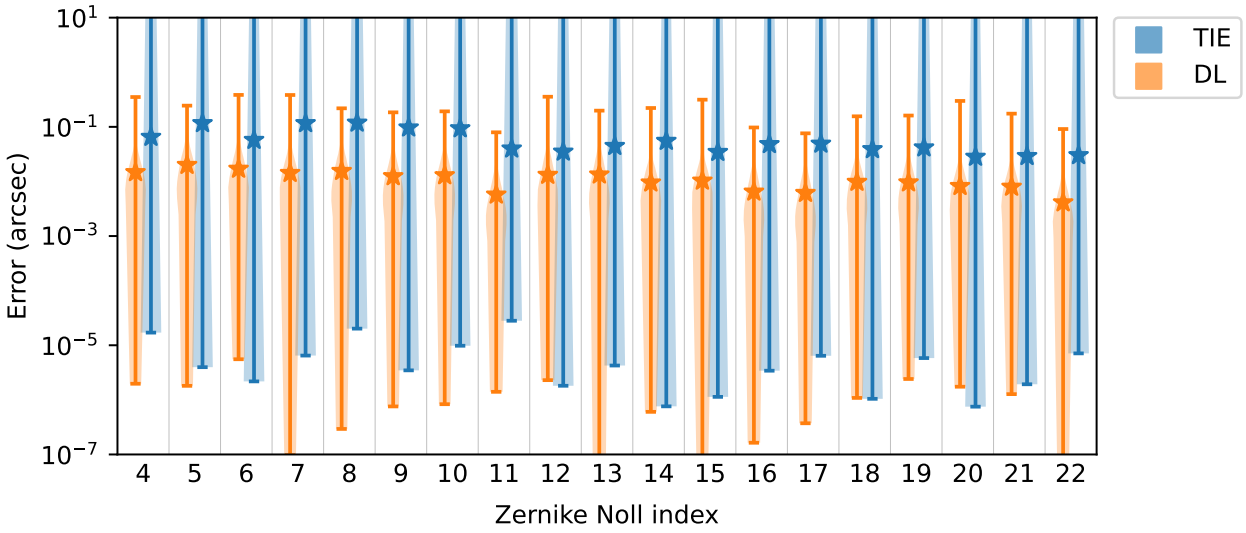

Figure C.21: Distribution of errors for each Zernike coefficient, α_i. The stars denote the median error.

Appendix D

DISENTANGLING SHAPE AND INTENSITY INFORMATION IN DEFOCUSED WAVE-FRONT ESTIMATION

Optical aberrations distort the PSF of a telescope. The Rubin Observatory AOS images stars out of focus to spread the PSF over a large number of pixels. This provides the spatial resolution necessary to diagnose the errors in the optical system.

In this chapter I demonstrate how this information manifests itself in out-of-focus images of stars that appear as donuts, with a particular focus on the differing information content provided by the shape and intensity patterns of these donuts. By studying how donut images are distorted by optical aberrations, we can understand the most salient information utilized by the AOS to estimate wave-front aberrations. This is useful for understanding the AOS, for diagnosing its performance, and for the kinds of manual, hands-on experimentation that takes place during telescope commissioning. This chapter complements papers from the literature, e.g., [153, 264], that examine the impact of optical aberrations on the PSF of in-focus images.

D.1 Aberrations in Defocused Images

In the absence of optical aberrations (and vignetting), defocused images of stars appear as axisymmetric "donuts" with radius

$$r_{\text{donut}} = \frac{1}{w_p} \frac{\ell}{\sqrt{4N^2 - 1}}, \tag{D.1}$$

where r_{donut} is in pixels, w_p is the pixel size, ℓ is the defocal offset, and N is the focal ratio of the system [68]. The radius of the "donut hole" is simply $r_{\text{hole}} = \varepsilon r_{\text{donut}}$, where ε is the fractional obscuration, i.e., the diameter of the central obscuration relative to the diameter of the pupil.

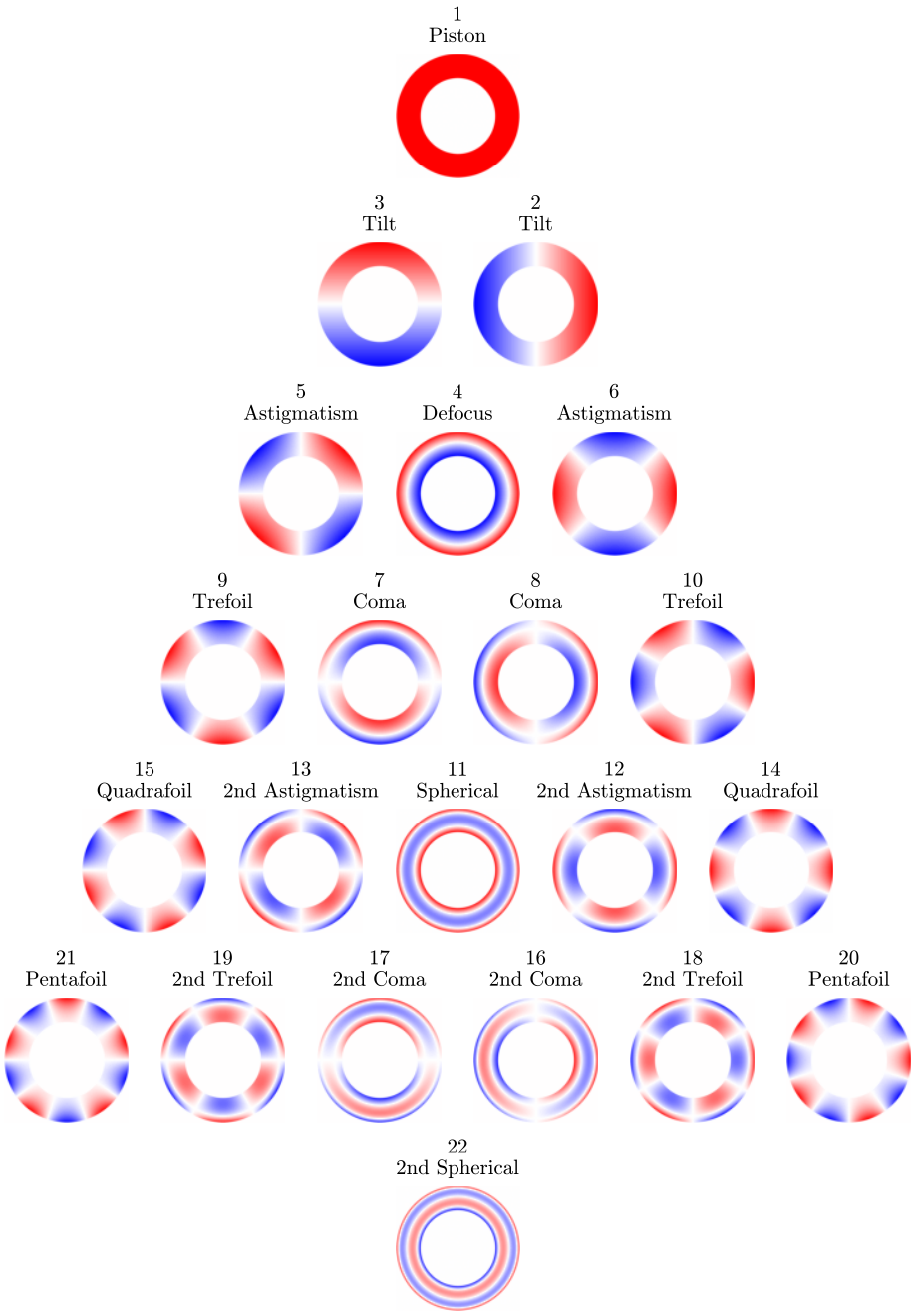

Figure D.1: Tree of Zernike polynomials for Noll indices through 22. Each is labeled with its Noll index and common aberration name. Rows are ordered by radial order, n, and columns are arranged by azimuthal order m.

j	ν	n	m	$R_n^m(\rho,\varepsilon)$	$Z_j(\rho,\theta;\varepsilon)$	Aberration
1	0	0	0	1	$R_0^0(\rho;\varepsilon)$	Piston
2	0	1	1	$\frac{\rho}{(1+\varepsilon^2)^{1/2}}$	$2R_1^1(\rho;\varepsilon)\cos(\theta)$	Tilt
3	0	1	1		$2R_1^1(\rho;\varepsilon)\sin(\theta)$	
4	0	2	0	$\frac{2\rho^2-1-\varepsilon^2}{1-\varepsilon^2}$	$\sqrt{3}R_2^0(\rho;\varepsilon)$	Defocus
5	0	2	2	$\frac{\rho^2}{(1+\varepsilon^2+\varepsilon^4)}^{1/2}$	$\sqrt{6}R_2^2(\rho;\varepsilon)\sin(2\theta)$	Astigmatism
6	0	2	2		$\sqrt{6}R_2^2(\rho;\varepsilon)\cos(2\theta)$	
7	1	3	1	$\frac{3(1+\varepsilon^2)\rho^3-2(1+\varepsilon^2+\varepsilon^4)\rho}{(1-\varepsilon^2)[(1+\varepsilon^2)(1+4\varepsilon^2+\varepsilon^4)]^{1/2}}$	$\sqrt{8}R_3^1(\rho;\varepsilon)\sin(\theta)$	Coma
8	1	3	1		$\sqrt{8}R_3^1(\rho;\varepsilon)\cos(\theta)$	
9	0	3	3	$\frac{\rho^3}{(1+\varepsilon^2+\varepsilon^4+\varepsilon^6)^{1/2}}$	$\sqrt{8}R_3^3(\rho;\varepsilon)\sin(3\theta)$	Trefoil
10	0	3	3		$\sqrt{8}R_3^3(\rho;\varepsilon)\cos(3\theta)$	
11	1	4	0	$\frac{6\rho^4-6(1+\varepsilon^2)\rho^2+1+4\varepsilon^2+\varepsilon^4}{(1-\varepsilon^2)^2}$	$\sqrt{5}R_4^0(\rho;\varepsilon)$	Spherical
12	1	4	2	$\frac{4\rho^4-3[(1-\varepsilon^8)/(1-\varepsilon^6)]\rho^2}{\{(1-\varepsilon^2)^{-1}[16(1-\varepsilon^{10})-15(1-\varepsilon^8)^2/(1-\varepsilon^6)]\}^{1/2}}$	$\sqrt{10}R_4^2(\rho;\varepsilon)\cos(2\theta)$	Secondary astigmatism
13	1	4	2		$\sqrt{10}R_4^2(\rho;\varepsilon)\sin(2\theta)$	
14	0	4	4	$\frac{\rho^4}{(1+\varepsilon^2+\varepsilon^4+\varepsilon^6+\varepsilon^8)^{1/2}}$	$\sqrt{10}R_4^4(\rho;\varepsilon)\cos(4\theta)$	Quadrafoil
15	0	4	4		$\sqrt{10}R_4^4(\rho;\varepsilon)\sin(4\theta)$	
16	2	5	1	$\frac{10(1+4\varepsilon^2+\varepsilon^4)\rho^5-12(1+4\varepsilon^2+4\varepsilon^4+\varepsilon^6)\rho^3+3(1+4\varepsilon^2+10\varepsilon^4+4\varepsilon^6+\varepsilon^8)\rho}{(1-\varepsilon^2)^2[(1+4\varepsilon^2+\varepsilon^4)(1+9\varepsilon^2+9\varepsilon^4+\varepsilon^6]^{1/2}}$	$\sqrt{12}R_5^1(\rho;\varepsilon)\cos(\theta)$	Secondary coma
17	2	5	1		$\sqrt{12}R_5^1(\rho;\varepsilon)\sin(\theta)$	
18	1	5	3	$\frac{5\rho^5-4[(1-\varepsilon^{10})/(1-\varepsilon^8)]\rho^3}{\{(1-\varepsilon^2)^{-1}[25(1-\varepsilon^{12})-24(1-\varepsilon^{10})^2/(1-\varepsilon^8)]\}^{1/2}}$	$\sqrt{12}R_5^3(\rho;\varepsilon)\cos(3\theta)$	Secondary trefoil
19	1	5	3		$\sqrt{12}R_5^3(\rho;\varepsilon)\sin(3\theta)$	
20	0	5	5	$\frac{\rho^5}{(1+\varepsilon^2+\varepsilon^4+\varepsilon^6+\varepsilon^8+\varepsilon^{10})^{1/2}}$	$\sqrt{12}R_5^5(\rho;\varepsilon)\cos(5\theta)$	Pentafoil
21	0	5	5		$\sqrt{12}R_5^5(\rho;\varepsilon)\sin(5\theta)$	
22	2	6	0	$\frac{20\rho^6-30(1+\varepsilon^2)\rho^4+12(1+3\varepsilon^2+\varepsilon^4)\rho^2-(1+9\varepsilon^2+9\varepsilon^4+\varepsilon^6)}{(1-\varepsilon^2)^3}$	$\sqrt{7}R_6^0(\rho;\varepsilon)$	Secondary spherical

Table D.1: Annular Zernike polynomials [193, 221]. j is the Noll index; ν is the number of times the radial polynomial turns over between $0 < \rho < 1$; n is the radial order; m is the azimuthal order; ε is the fractional obscuration.

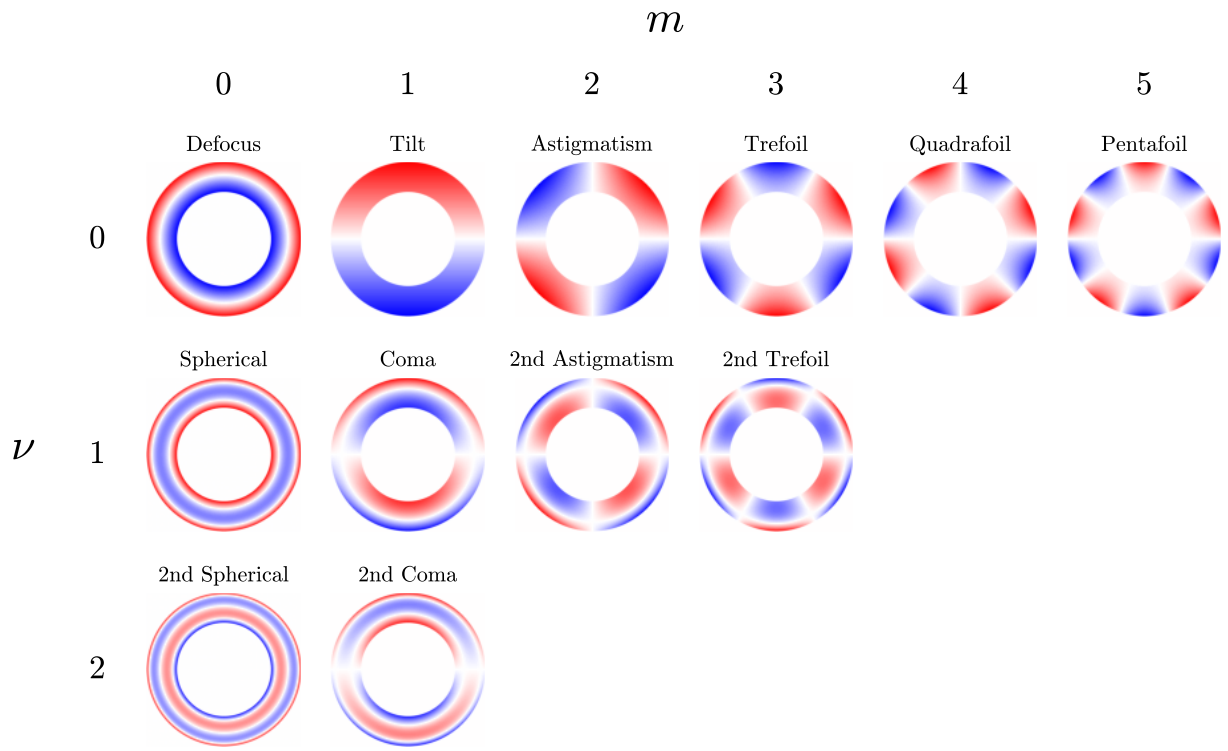

Figure D.2: Zernikes polynomials for Noll indices 2-22. Rows are grouped according to ν, the number of times the radial polynomial turns over between $0 < \rho < 1$. Columns are arranged according to the azimuthal order, m. To conserve space, only the $\sin(m\theta)$ modes are shown for $m \geq 1$.

The size, shape, and intensity of these donuts change in the presence of optical aberrations. We express these aberrations in terms of annular Zernike polynomials (Table D.1, Figure D.1), a complete orthonormal basis on the annulus. Each Zernike polynomial is closely associated with one of the classical optical aberrations [264], whose names are listed in the far right column of Table D.1. In this section we study how each of these aberrations manifest in the shape and intensity variations of defocused donut images. We will find that it is useful to group aberrations by ν, the number of times the radial polynomial turns over. Figure D.2 rearranges the Zernike polynomials from Figure D.1 by ν and m.

To characterize the signal associated with each aberration, we examine the photon displacements generated by the wavefront:

$$\Delta\boldsymbol{\rho} \propto -\nabla W(\boldsymbol{\rho}). \tag{D.2}$$

This provides an analytic handle on the shape distortions generated by each Zernike mode. Additionally, we examine the divergence of these displacements (equivalent to the negative wavefront curvature) to determine whether each aberration generates surface brightness fluctuations in the defocused donuts:

$$\nabla \cdot \Delta\boldsymbol{\rho} \propto -\nabla^2 W(\boldsymbol{\rho}). \tag{D.3}$$

In particular, variations in surface brightness are generated by wavefront curvature with non-trivial coordinate dependence.

In addition we use two figures to visualize how each aberrations impacts defocused donuts. The first is the projection from the pupil to the image plane of a set of concentric circles with radii $\varepsilon < \rho < 1$. In the absence of any aberrations, these projections will remain perfectly circular and concentric on the image plane. Examining how these projections are distorted and translated relative to each other provides insight into the shape distortions induced by each aberrations, as well as understanding how surface brightness fluctuations are generated. The second is a simulated image with point spread function PSF FWHM of 0.5".

Without loss of generality, for aberrations with $m \geq 1$, we will consider only the $\sin(m\theta)$ modes. The $\cos(m\theta)$ modes are merely rotations of the $\sin(m\theta)$ modes, and all of the conclusions about the latter apply to the former. Furthermore, we will only plot the impact on intrafocal donuts. The impact on extrafocal donuts is equivalent to a sign flip in the wavefront.

D.1.1 $\nu = 1$ aberrations

The $\nu = 1$ aberrations are monotonic in ρ. We will find that all of these modes change the shape of the pupil without generating fluctuations in surface brightness.

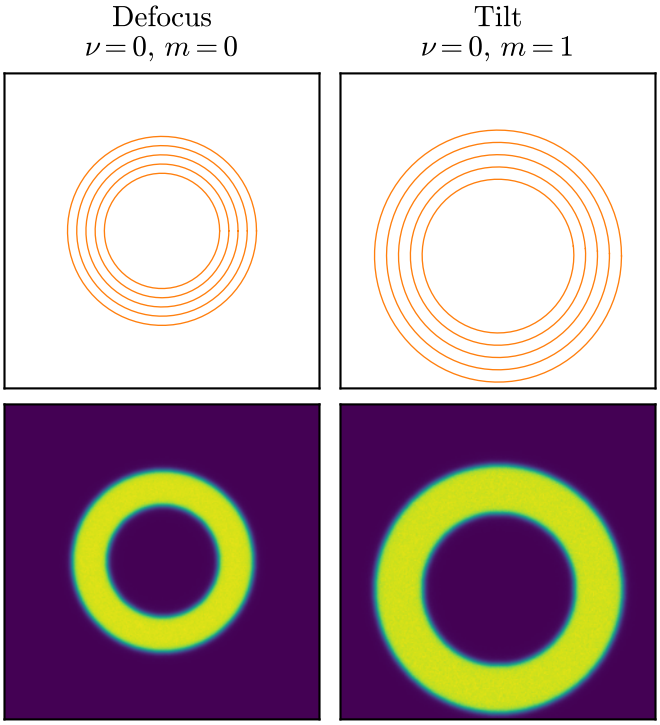

Figure D.3: Positive defocus and tilt applied to intrafocal donuts. Top: mapping of concentric circles from the pupil to the image plane. Bottom: simulated intrafocal images.

Defocus ($\nu = 1,\ m = 0$)

A wavefront consisting of pure defocus scales as

$$W(\rho, \theta) \propto \rho^2 + c_{2,0}(\varepsilon), \tag{D.4}$$

where $c_{2,0}(\varepsilon)$ is a constant that depends on the fractional obscuration, ε. The photon displacement therefore scales as

$$\Delta\boldsymbol{\rho} \propto -\rho\,\hat{\boldsymbol{\rho}}, \tag{D.5}$$

from which one can see that positive defocus shrinks intra-focal images and stretches extra-focal images, while preserving concentric circles. This can be seen in the upper-left panel of Figure D.3 (cf. the size of the circles in the top-right panel).

Importantly, the curvature of this mode is constant:

$$\nabla^2 W(\boldsymbol{\rho}) = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial W}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 W}{\partial \theta^2} = \text{const.}, \tag{D.6}$$

such that the shrinking/stretching of defocus scales the donut surface brightness uniformly. In other words, under positive (negative) defocus, each point on the intrafocal donut acts as a sink (source), and vice versa for extrafocal donuts. There is, therefore, no information about defocus in surface brightness fluctuations. This can be seen from the uniformity of the donut image in the bottom-left panel of Figure D.3.

Tilt ($\nu = 1$, $m = 1$)

Tilt merely translates donuts, as we can clearly see from the gradient of a pure-tilt in Cartesian coordinates:

$$W(\boldsymbol{\rho}) \propto \rho \sin\theta = y \quad \rightarrow \quad \Delta\boldsymbol{\rho} \propto -\hat{\boldsymbol{y}}. \tag{D.7}$$

In particular, because there is no ρ or θ dependence, the entire donut is shifted as if it were a rigid object, as seen in the right two panels of Figure D.3. Tilt, therefore, leaves no discernible imprint on individual donut images, and is therefore typically ignored during wave-front estimation.

Astigmatism and m-foil ($\nu = 1$, $m \geq 2$)

Astigmatism and m-foil aberrations (e.g., trefoil, quadrafoil, pentafoil, etc.) have wavefronts that scale as

$$W(\boldsymbol{\rho}) \propto \rho^m \sin(m\theta), \tag{D.8}$$

generating photon displacements proportional to

$$\Delta\boldsymbol{\rho} \propto -m\,\rho^{m-1}\big[\,\sin(m\theta)\hat{\boldsymbol{\rho}} + \cos(m\theta)\hat{\boldsymbol{\theta}}\,\big]. \tag{D.9}$$

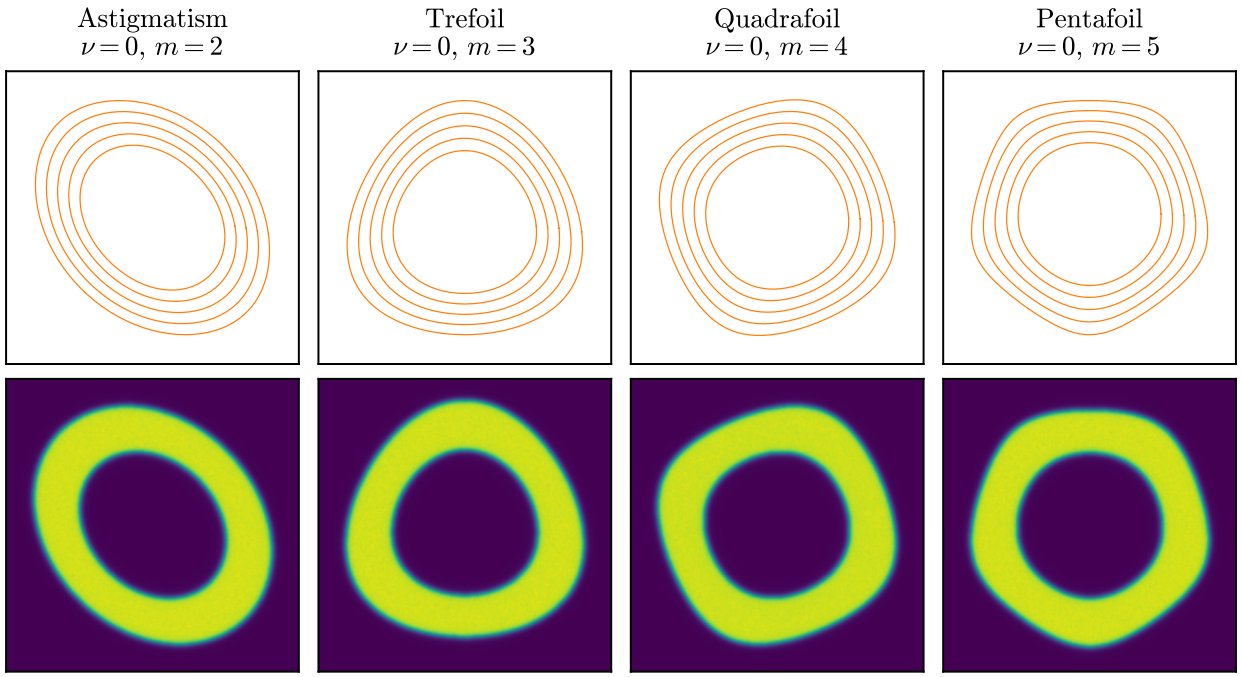

Figure D.4: Positive astigmatism, trefoil, quadrafoil, and pentafoil applied to intrafocal donuts. Top: mapping of concentric circles from the pupil to the image plane. Bottom: simulated intrafocal images. Note that, for illustrative purposes, the amplitude of these modes is much greater than typically expected in regular telescope operation.

The $\hat{\boldsymbol{\rho}}$ term in this displacement generates a standing wave with m maxima. The $\hat{\boldsymbol{\theta}}$ term pushes points towards (away from) the maxima (minima) of the $\hat{\boldsymbol{\rho}}$ term, thereby making the standing waves "sharper". The top row of Figure D.4 shows how each mode distorts concentric circles to resemble a rounded m-sided polygon.

For all modes of form $W(\boldsymbol{\rho}) \propto \rho^m \sin(m\theta)$, it is straightforward to show that $\nabla^2 W(\boldsymbol{\rho}) = 0$. Thus primary astigmatism and m-foil modes, like the other $\nu = 0$ modes, do not generate any surface brightness fluctuations. This is evident in the bottom row of Figure D.4.

Finally, note that a negative amplitude results in a rotation by π/m:

$$-\sin(m\theta) = \sin(m(\theta + \pi)) = \sin(m\theta + \frac{\pi}{m}). \tag{D.10}$$

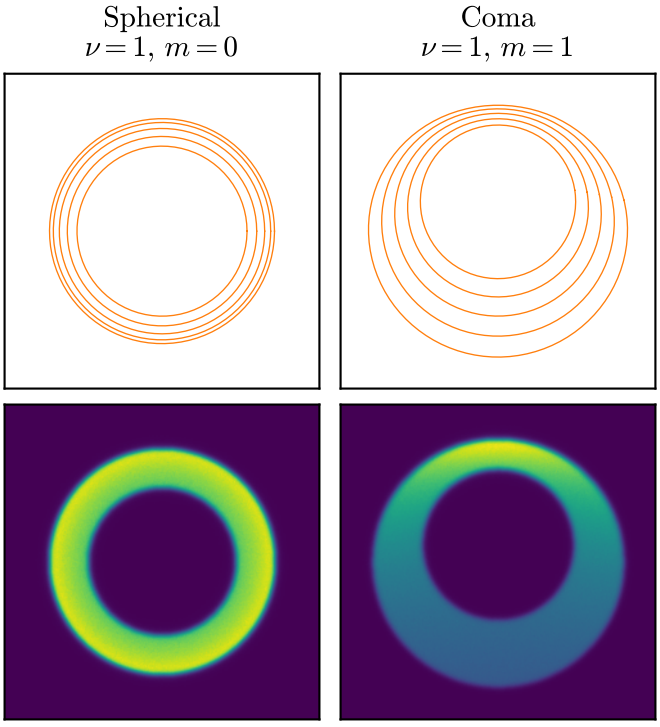

Figure D.5: Positive spherical aberration and coma applied to intrafocal donuts. Top: mapping of concentric circles from the pupil to the image plane. Bottom: simulated intrafocal images.

This corresponds to reflecting the donuts in Figure D.4 about the horizontal axis.

D.1.2 $\nu = 2$ aberrations

The radial functions for $\nu = 2$ turn over once at $\rho < 1$. We will find that these aberrations change the donut shape in similar ways to the $\nu = 0$ aberrations, however the $\nu = 2$ aberrations reshape the outer edge far more than the inner edge. This necessitates non-trivial changes to the surface brightness in the donut interiors.

Spherical ($\nu = 2$, $m = 0$)

A wavefront consisting of pure spherical aberration scales like

$$W(\rho, \theta) \propto \rho^4 - (1 + \varepsilon^2)\rho^2 + c_{4,0}(\varepsilon), \tag{D.11}$$

where $c_{4,0}(\varepsilon)$ is a constant that depends on the fractional obscuration, ε. The gradient of this wavefront generates photon displacements

$$\Delta\boldsymbol{\rho} \propto \big[-2\rho^3 + (1 + \varepsilon^2)\rho \big] \hat{\boldsymbol{\rho}}, \tag{D.12}$$

from which it is easy to see that spherical aberration, like defocus, preserves concentric circles. However, these circles, unlike defocus, are not all displaced in the same direction. In particular,

$$\Delta\rho\,|_{\rho=1} < 0 \quad \text{while} \quad \Delta\rho\,|_{\rho=\varepsilon} \propto \varepsilon(1 - \varepsilon^2) > 0, \tag{D.13}$$

for any $0 < \varepsilon < 1$. The stationary midpoint corresponds to the circle of radius

$$\rho = \sqrt{\frac{1 + \varepsilon^2}{2}}. \tag{D.14}$$

Interestingly, this splits the donut into inner and outer halves by area.

Positive (negative) spherical aberration, therefore, pushes the inner and outer edges towards (away from) each other, thereby making the donut look skinnier (fatter). This can be seen in the top-left panel of Figure D.5, where an intrafocal donut is made skinnier by positive spherical aberration. The opposite effect occurs for extrafocal donuts.

Compressing or dilating the donut necessarily results in a non-zero wave-front curvature. Unlike defocus, however, the curvature of spherical modes has a non-trivial dependence on ρ:

$$\nabla^2 W(\boldsymbol{\rho}) \propto 4\rho^2 - (1 + \varepsilon^2), \tag{D.15}$$

indicating that spherical aberration generates intensity fluctuations in the donut. This is hinted at by the uneven line spacing in the top left panel of Figure D.5, and is seen explicitly in the bottom-left panel of the same figure.

Note that while spherical aberration produces non-trivial surface brightness fluctuations, it also produces a distinctive impact on the shape of the donut (making it skinnier or fatter). In principle, therefore, this mode can be detected from shape information alone.

Coma ($\nu = 2$, $m = 1$)

Coma is easiest to understand in Cartesian coordinates:

$$W(x, y) \propto \big[3(1 + \varepsilon^2)(x^2 + y^2) - 2(1 + \varepsilon^2 + \varepsilon^4)\big] y, \tag{D.16}$$

for which the photon displacement, up to a constant shift, is proportional to

$$\begin{aligned} \Delta\boldsymbol{\rho} \propto 3(1 + \varepsilon^2)\rho^2 \big[-\sin 2\theta\, \hat{\boldsymbol{x}} + \cos 2\theta\, \hat{\boldsymbol{y}} \big] \\ -6(1 + \varepsilon^2)\rho^2 \hat{\boldsymbol{y}}. \end{aligned} \tag{D.17}$$

The first term in this equation slightly squashes concentric circles, however this effect is dominated by the second term which displaces each circle by a distance proportional to the square of its radius. This latter term is responsible for the "comet tail" traditionally associated with coma. In defocused donut images, it is responsible for displacing the "donut hole" relative to the outer donut edge, as seen in the right panels of Figure D.5.

The relative shift of the inner and outer donut edges necessarily generates surface brightness fluctuations. Indeed it is straightforward to verify that the wavefront curvature bears non-trivial radial and azimuthal dependencies. These fluctuations can be seen in the right panels of Figure D.5 where it is clear the fluctuations in surface brightness are due to the compression and rarefaction of the donut surface along the comatic axis.

Note that while coma produces non-trivial surface brightness fluctuations, it also produces a distinctive impact on the shape of the donut (mis-centering the donut hole). In principle, therefore, this mode can be detected from shape information alone.

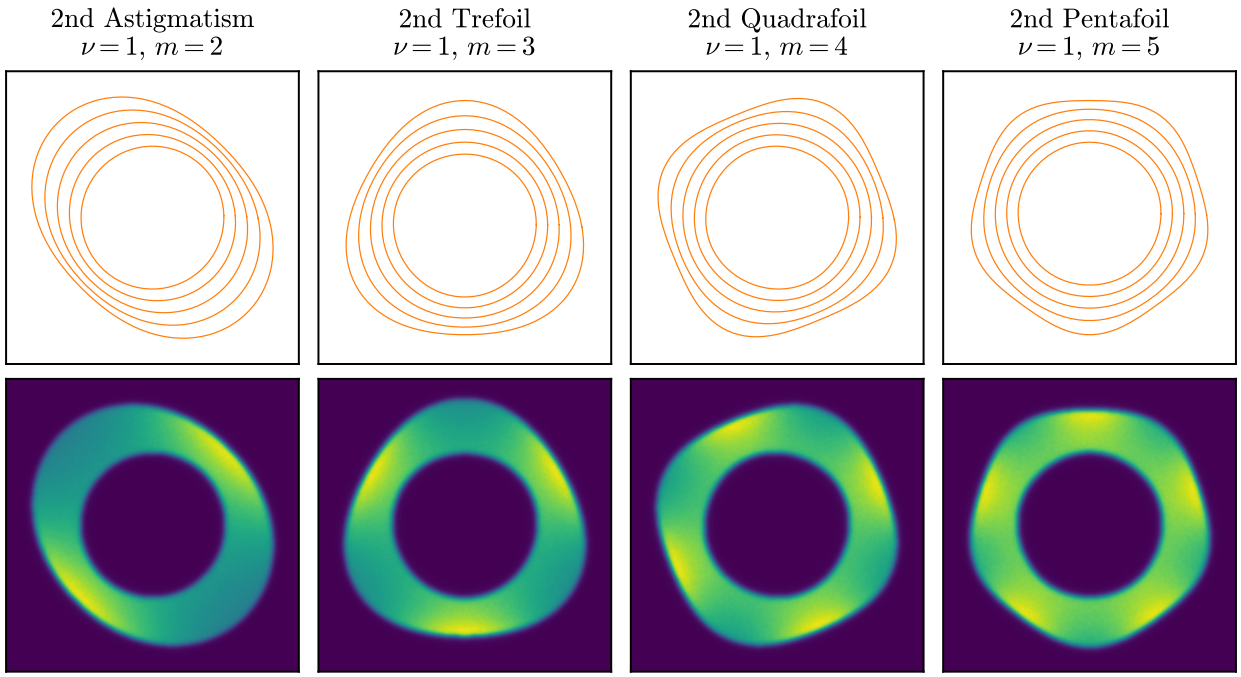

Figure D.6: Positive secondary astigmatism, trefoil, quadrafoil, and pentafoil applied to intrafocal donuts. Top: mapping of concentric circles from the pupil to the image plane. Bottom: simulated intrafocal images. Note that, for illustrative purposes, the amplitude of these modes is much greater than typically expected in regular telescope operation.

Secondary astigmatism and m-foil ($\nu = 2,\ m \geq 2$)

Secondary astigmatism and m-foil aberrations have wavefronts that scale approximately as

$$W(\boldsymbol{\rho}) \propto (2m\rho^{2m} - (m+1)\rho^{m}) \sin(m\theta), \tag{D.18}$$

where in the second term we have suppressed a factor of

$$\frac{1-\varepsilon^{2(m+2)}}{1-\varepsilon^{2(m+1)}} \sim 1. \tag{D.19}$$

This wavefront generates a photon displacement

$$\begin{aligned}\Delta\boldsymbol{\rho} \propto\ & 2m\rho^{2m-1}\big[\,2\sin(m\theta)\hat{\boldsymbol{\rho}} + \cos(m\theta)\hat{\boldsymbol{\theta}}\,\big] \\ & - (m+1)\rho^{m-1}\big[\,\sin(m\theta)\hat{\boldsymbol{\rho}} + \cos(m\theta)\hat{\boldsymbol{\theta}}\,\big].\end{aligned} \tag{D.20}$$

The second term of this equation is nearly identical to the primary astigmatism and m-foil displacement (Eq. D.9), with an opposite sign. The first term, while superficially different, produces nearly the same effect. Together, the displacement of secondary astigmatism and m-foil aberrations is very similar to that of the primary aberrations. The larger radial exponent in the first term of Eq. D.20, however, ensures that the outer donut edge is more distorted than the inner edge. In other words, the secondary aberrations produce similar distortion of the outer edge, while leaving the inner edge nearly circular.

Furthermore, unlike the primary aberrations, secondary astigmatism and m-foil produces non-trivial fluctuations in surface brightness. This can be verified by checking that the wavefront curvature of these modes have non-trivial radial and azimuthal dependencies, and by noting the intensity pattern visible in Figure D.6.

Note that while secondary astigmatism and m-foil produce non-trivial surface brightness fluctuations, they also produce distinctive impacts on the shape of the donut. In particular, they reshape the outer edge while leaving the inner edge nearly circular, while the primary modes reshape the inner edge to a greater degree. By comparing Figure D.4 and Figure D.6, however, one sees this distinction between the primary and secondary modes becomes less pronounced for larger m. This is because the radial scaling of the primary mode is ρ^{m-1}, which is already steep for large m, and because polygons with more sides are already closer to circular.

In principle, therefore, lower-m secondary modes (e.g, astigmatism and trefoil) can be distinguished from the corresponding primary modes on the basis of shape information alone. On the other hand, distinguishing primary and secondary aberrations for higher-m modes must rely on surface brightness fluctuations.

D.1.3 $\nu = 3$ aberrations

The radial functions for $\nu = 3$ aberrations turn over twice in the domain $r < 1$. We consider only two of these aberrations: secondary spherical ($\nu = 3$, $m = 0$) and secondary coma ($\nu = 3$, $m = 1$). We do not pursue an analytic analysis here, as the results are cumbersome.

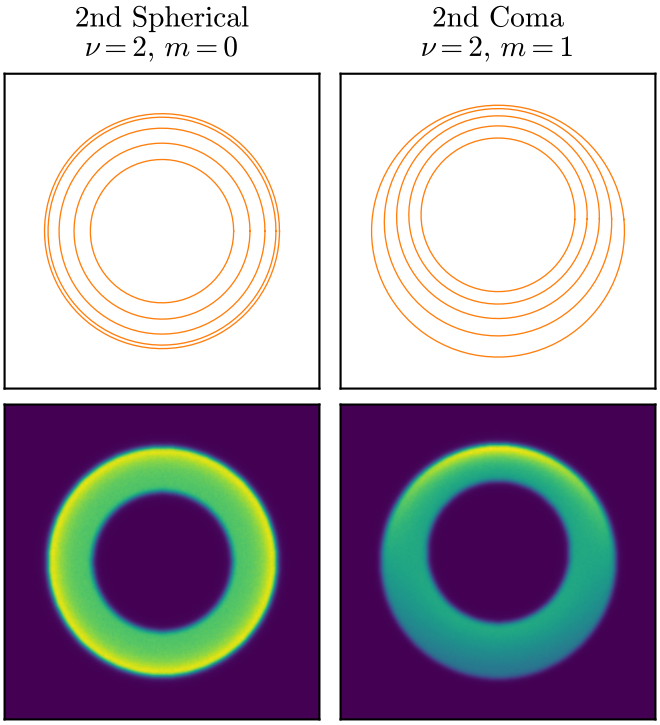

Figure D.7: Positive secondary spherical aberration and coma applied to intrafocal donuts. Top: mapping of concentric circles from the pupil to the image plane. Bottom: simulated intrafocal images.

We note, however, that these aberrations continue the trends we have seen before: higher-ν aberrations generate increasingly non-trivial surface brightness fluctuations, with the impact shifting closer towards $\rho = 1$.

The impacts of secondary spherical and coma aberrations are shown in Figure D.7. In every panel of this figure one can see that the inner edge of the donut is less perturbed than the corresponding panels in Figure D.5. In addition, the bright regions generated from the compression of concentric circles shifts farther towards the outer rim of the donut.

Note these modes have a very small impact on the shape of the donut, even as they approach the caustic limit (as the amplitudes shown in Figure D.7 are). As such, one must rely on surface brightness fluctuations to detect these aberrations.

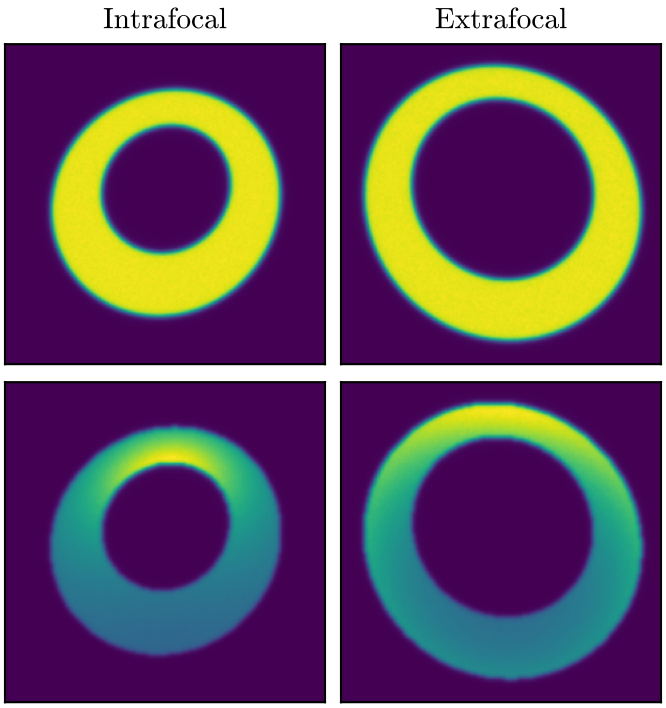

Figure D.8: Example pair of intra- and extra-focal donuts (left and right, respectively). The top row presents only shape information, while the bottom row contains surface brightness fluctuations. For analysis of these donuts, see Section D.2.

D.2 Example of By-Eye Wave-front Estimation

In this section we provide a worked example of by-eye wavefront estimation using the principles described in the previous section. This demonstrates how real-time diagnostics and sanity checks are performed while commissioning an active optics system, during which commissioning scientists are iteratively taking images, examining defocused donuts, making inferences about the state of the optical system, and comparing these inferences to the estimates and behavior of the active optics system. We encourage the reader to attempt this exercise themselves before reading the analysis below.

Figure D.8 shows a simulated intra- and extra-focal donut pair. The top row displays only shape information, while the bottom row contains surface brightness fluctuations. There are

a number of inferences that can be made from the shape of the donuts alone:

- The intrafocal donut is smaller than the extrafocal donut, indicating a positive defocus (Z4).

- The intrafocal donut is slightly fatter than the extrafocal donut (cf. the bottom rim, which looks slightly thicker for the intrafocal donut, even though this donut is smaller), indicating a negative spherical aberration (Z11).

- The donut hole is decentered towards the top of the image, indicating positive coma along the y axis (Z7).

- The donuts are oblong along opposite diagonal axes, with the intrafocal donut tilted right, indicating negative oblique astigmatism (Z5).

There is one further piece of information we can gain from the surface brightness fluctuations:

- The bright spot on the comatic axis of the intrafocal donut is pushed *away* from the exterior of the donut, which may indicate negative secondary coma along the y axis (Z17). Note, however, this effect can also be generated by negative spherical aberration (Z11), which we do know is present from the shape (see above).

The inferences above demonstrate how significant information can be gleaned about lower-order aberrations from the shape of the donuts alone. Surface brightness fluctuations, on the other hand, can provide information about higher-order modes. Caution is warranted, however, as there exist surface-brightness degeneracies between modes when considering combinations of multiple aberrations, more so than when considering donut shapes. Indeed, in the example above, the surface brightness fluctuation used to infer the presence of secondary coma is in fact generated by a combination of secondary coma and spherical aberration. In addition, we do not usually attempt to infer and correct all of these aberrations at once, but rather correct the largest lower-order aberrations first. We also typically attribute significant

features to lower-order modes, as these are far more likley to be excited at the level required for the changes seen.

With knowledge of how optical misalignment and mirror surface perturbations generate different aberrations, one can use the inferences above to propose and test corrections to the optical system. For example, coma and astigmatism are often generated by decentering and tilting the secondary mirror, respectively. For more discussion of these relationships, see [199, 264].

D.3 Investigations with Algorithmic Wave-front Estimation

In this section we use the Danish wave-front estimator to explore the nuances of shape vs intensity information. Danish forward models donut images for a given telescope geometry and set of aberrations, convolves the image with an atmospheric kernel, and fits the aberration Zernike coefficients by minimizing residuals with the true donut image. See R. J. Janish [152] and the previous chapter for more information.

We compare results from Danish to results from a modified version of Danish that only predicts donut shapes. The modified version uses the same framework as regular Danish, but sets all donut pixels to the same values (except for pixels on the edge of the donut, which are illuminated according to the fraction of the pixel covered by the donut).

D.3.1 Impact of Zernike sparsity

First we compare the effectiveness of shape-only wavefront estimation to full-information (shape + intensity) estimation using only the lowest-order ν modes. Recall that piston and tilt ($\nu = 0; m = 0, 1$) are not detectable in wave-front estimation, so for $m = 0, 1$, we estimate the $\nu = 1$ modes (spherical aberration and coma). The results in Figure D.9 demonstrates these modes can be accurately estimated using only the shape of the donuts, and the intensity information provides almost no additional information. This is an empirical confirmation of the conclusions in the previous section.

Next we perform the same comparison using all Zernike coefficients between Noll indices

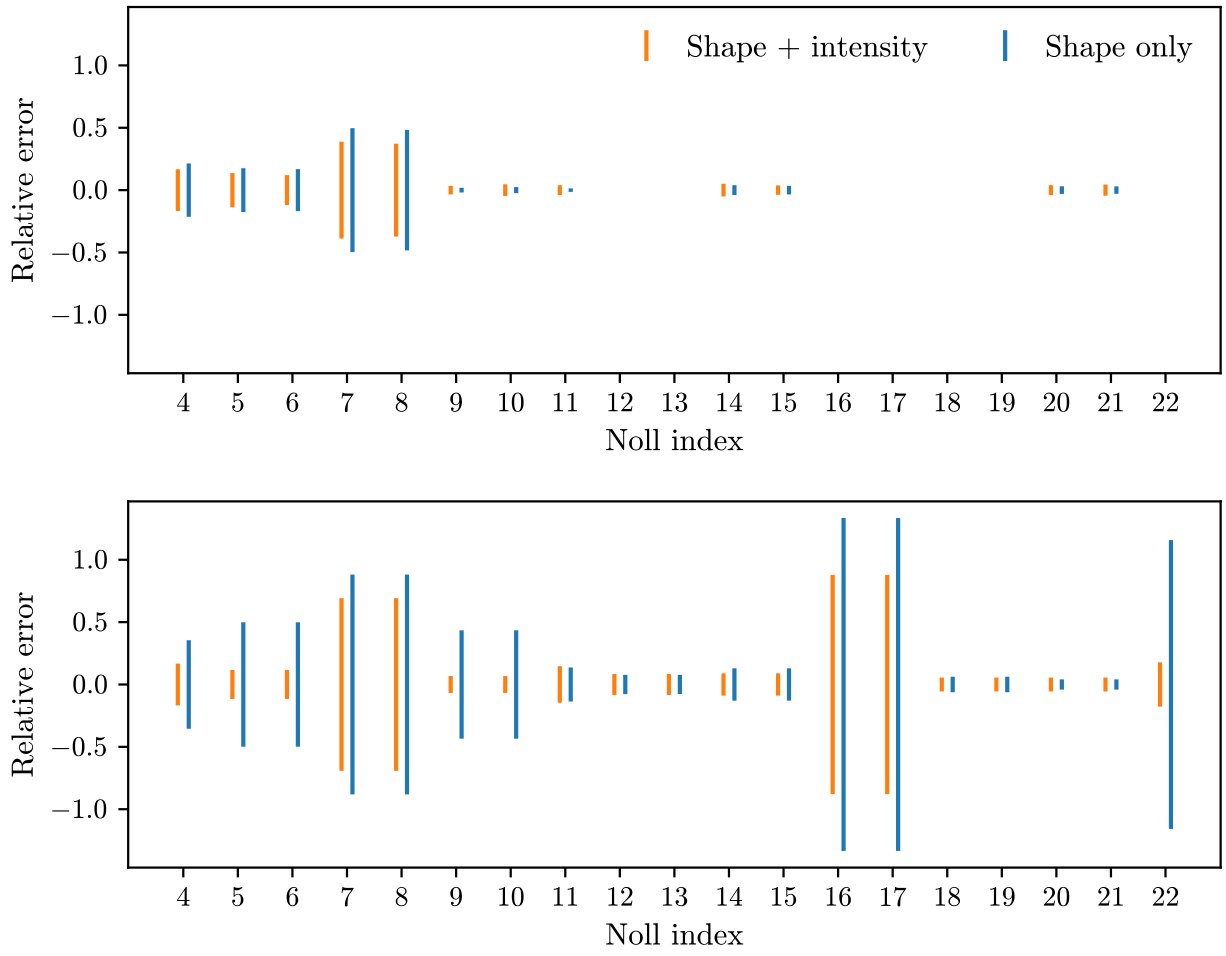

Figure D.9: Comparing wave-front estimation errors using only shape information vs shape + intensity information. Top: "sparse" Zernike estimation, where we only estimate those coefficients that have distinctive impacts on the shape of the donut. Bottom: "dense" Zernike estimation, including all coefficients between Noll indices 4-22. All lines are 1σ error bars, centered on zero.

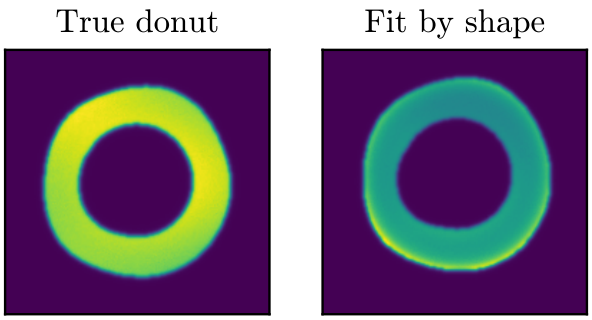

Figure D.10: Example of shape degeneracy in wave-front estimation. Left: true donut shape and intensity. Right: model donut, fit by shape only. These two donuts have similar shapes but very different intensity patterns, as the result of higher-order aberrations. Due to these degeneracies, it is dangerous to fit higher-order aberrations using only shape information.

4-22 (i.e. all aberrations discussed in this chapter). In this case we see there is significant information present in the intensity information. Very few aberrations are well fit using only shape information, including the $\nu = 0$ modes. This is because, without intensity information, there are substantial degeneracies between higher- and lower-order ν modes. Note that even when including intensity information Danish struggles to estimate first- and second-order coma. This is a known difficulty that is further investigated in the next section.

Note that using the donut shape to infer lower-order ν modes makes the assumption that the higher-order ν modes are not important. It is true, however, that when higher-order modes are excited, there exist significant degeneracies in the shape of the donut. This is shown in Figure D.10, in which Danish is used to fit the wavefront of the donut in the left panel using only shape information. The fit wavefront is then used to predict the donut image, which is shown on the right. You can see the shapes are very similar, but the intensity patterns are very different, with the donut on the right approaching the caustic regime on its lower edge.

This demonstrates how, when looking at the donut shape to infer the presence of lower-

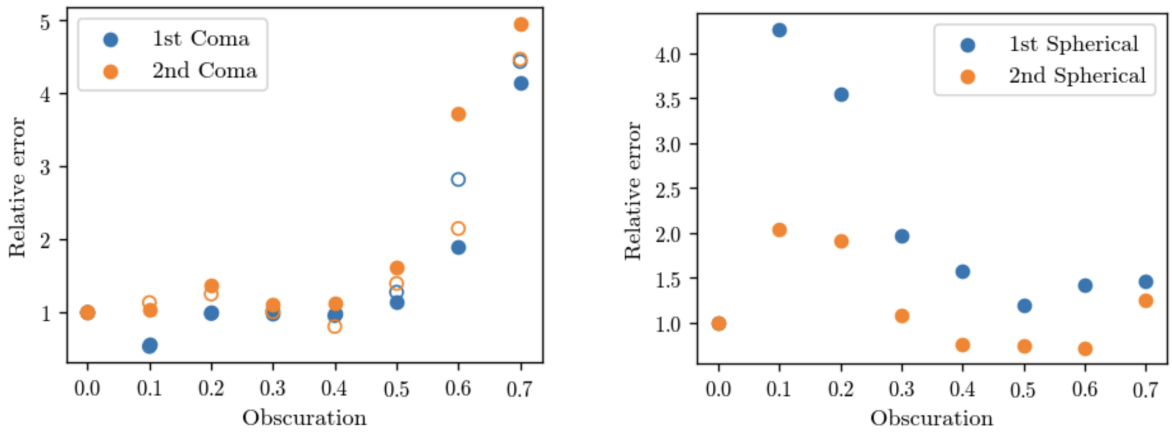

Figure D.11: The impact of the central obscuration on wavefront estimation. Left: increasing the size of the central obscuration degrades coma estimation. Filled and hollow points represent first and second coma of different parity. Right: increasing the size of the central obscuration improves estimation of primary spherical aberration, while secondary spherical estimation is mostly unaffected.

order ν modes, the donut intensity implicitly plays an important role by placing a prior on the higher-order modes. In particular, when the donut intensity pattern is smooth, it is usually fair to assume the donut shape is dominated by lower-order ν modes rather than extreme combinations of higher-order ν modes.

D.3.2 Impact of the central obscuration

There is one additional aspect of the donut shape that impacts wave-front estimation: the size of the central obscuration. The central obscuration for Rubin is quite large, effectively removing a large part of the central wavefront, making wavefront estimation more difficult by removing much of the intensity information that could be used to disentangle first order and second order aberrations (e.g., coma vs 2nd coma).

We investigate the impact of the central obscuration by simulating donut images while

varying the size of the obscuration, and estimating Zernike coefficients to compare the relative error. Results are shown in Figure D.11. The left panel demonstrates the behavior exhibited by most $\nu > 0$ aberrations: that increasing the size of the central obscuration has very little impact for $\varepsilon < 0.5$, however increasing the obscuration beyond this point quickly degrades the accuracy of Zernike estimation. Unfortunately, Rubin has an obscuration of $\varepsilon = 0.61$, putting us in the degraded estimation regime[1].

Spherical aberration, however, demonstrates the opposite behavior: it is quite difficult to estimate spherical aberration when the obscuration is small. The accuracy of spherical aberation estimation increases with the size of the central obscuration until $\varepsilon \sim 0.5$. This is intuitive as the only sign of spherical aberration is the opposite movements of the inner and outer donut edges. As the central obscuration gets smaller, it's motion under spherical aberration approaches zero and the inner edge is measured by fewer pixels. The error associated with secondary spherical aberration, however, is fairly stable under changes to the central obscuration, as secondary spherical primarily changes the donut intensity near the outer edge (cf. Figure D.7).

D.4 Conclusion

In this chapter we used analytic and simulation tools to investigate how shape and intensity information are used in defocused wavefront estimation. We introduced a new ordering parameter, ν, which corresponds to the number of times the radial polynomial turns over. We find that aberrations with $\nu = 0$ impact on the shape of the donut, and leave no signature in the intensity pattern of the donuts. As ν increases the impact on donuts shifts progressively away from modifying the shape towards changing the intensity pattern in the interior of the donut.

We show that the lowest-order ν modes can be well characterized using only donut shapes, while higher-order modes require combining shape and intensity information. While using

[1]Note this is not truly unfortunate, as the large central obscuration is a central design component that enables Rubin's unique wide-and-deep imaging capabilities.

shape information to infer the amplitude of lower-order ν modes, however, the intensity information is still important in that it acts as a prior on the higher-order ν modes. Specifically, if the intensity patterns in the donut are smooth and show no signs of caustics, it is usually a good assumption that prominent donut features are dominated by the lower-order ν modes.

Finally, we show that the difficulty associated with estimating higher-order aberrations in the Rubin Observatory, even when using intensity information, is due to the large central obscuration of the Rubin system. This is except for spherical aberration, which is easier to estimate with a large obscuration due to the fact that spherical aberration moves the inner and outer donut edges in opposite directions with a magnitude that increases with increasing obscuration size.

We believe the findings of this chapter will be helpful for building intuition for wavefront estimation and aid scientists that find themselves looking at donut images while commissioning and experimenting with active optics systems on any telescope that uses defocused wavefront sensing.